US 20070092943A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0092943 A1**

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

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

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

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

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

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

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

Related U.S. Application Data

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

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

(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.

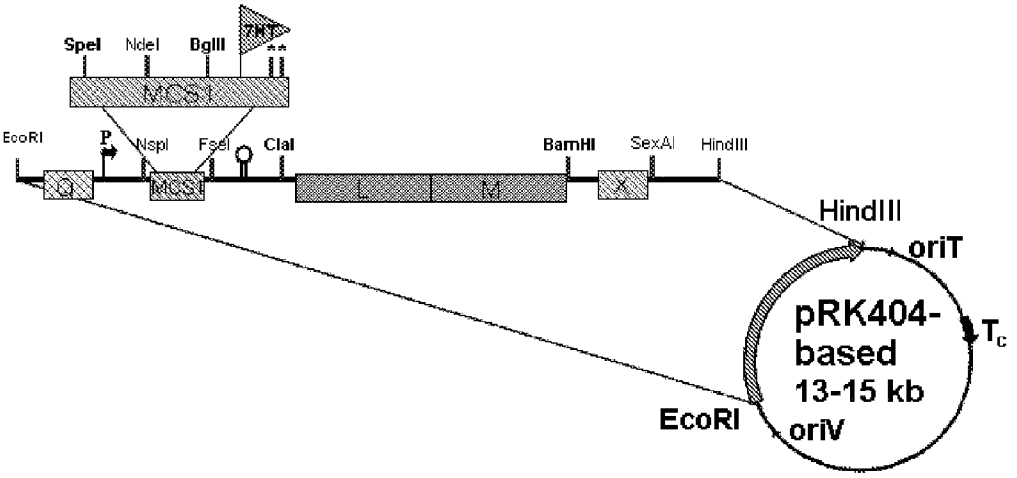
SpeI
NdeI
BglII
7HT
MCS1
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS2
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 1

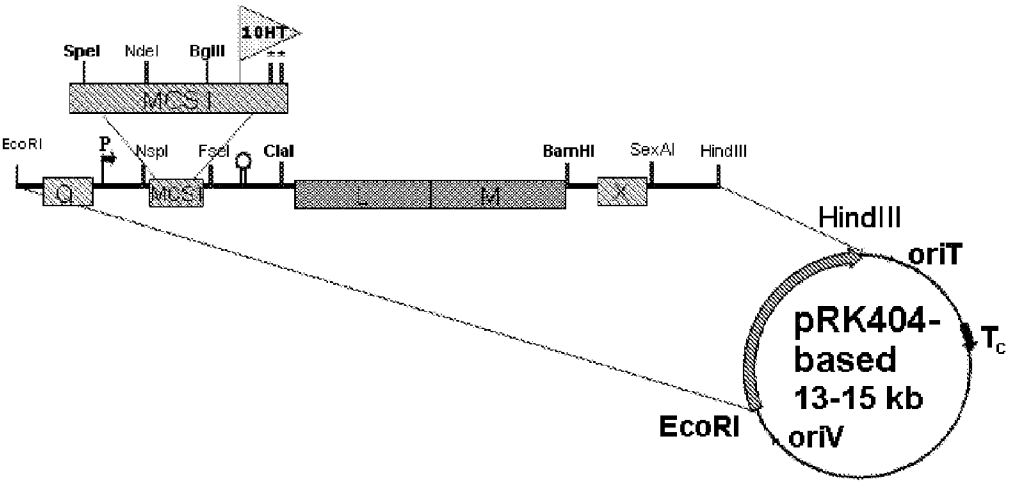
SpeI
NdeI
BglII
10HT
MCS1
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS2
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 2

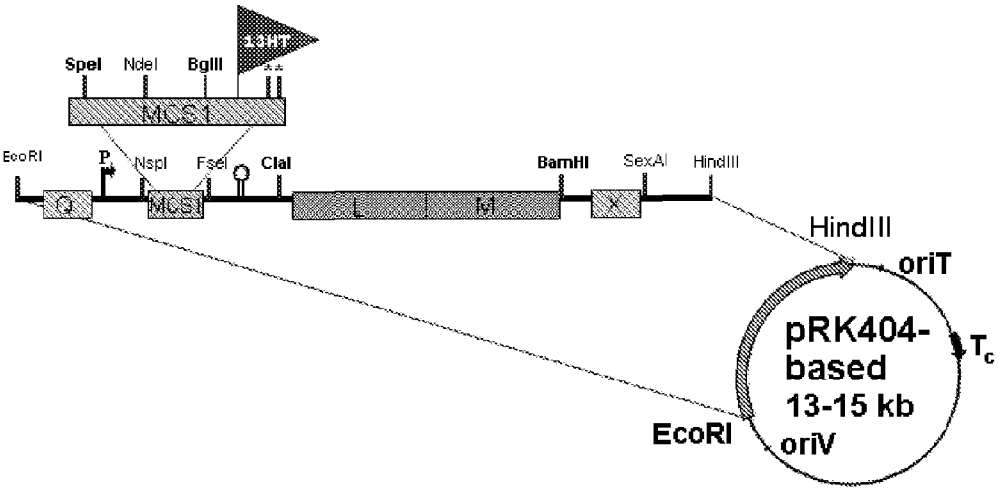
13HT
SpeI
NdeI
BglII
MCS1
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS1
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
T_c
EcoRI

FIG. 3

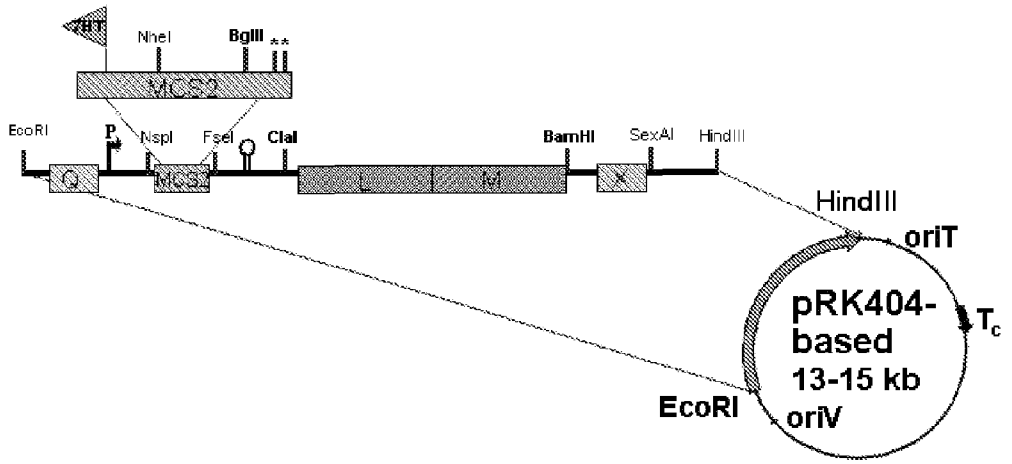
7HT
NheI
BglII
MCS2
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS2
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
T_c
EcoRI

FIG. 4

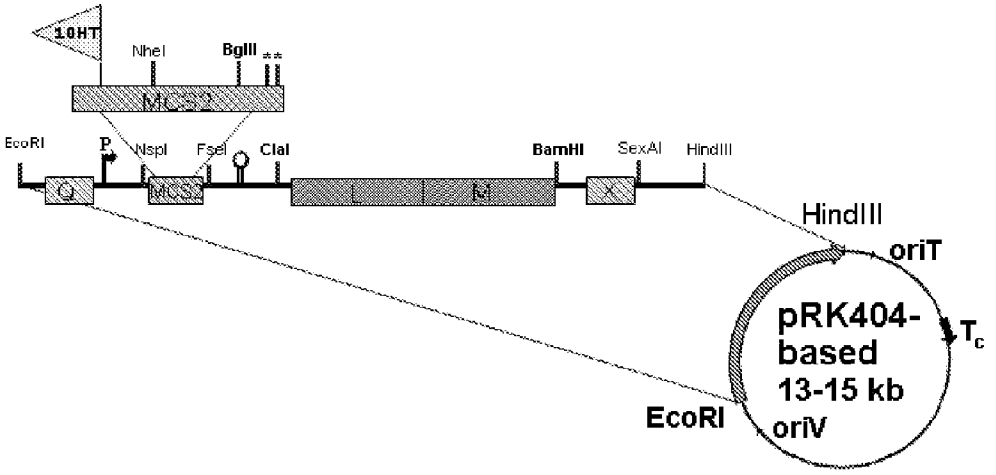
10HT
NheI
BglII
MCS2
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS1
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 5

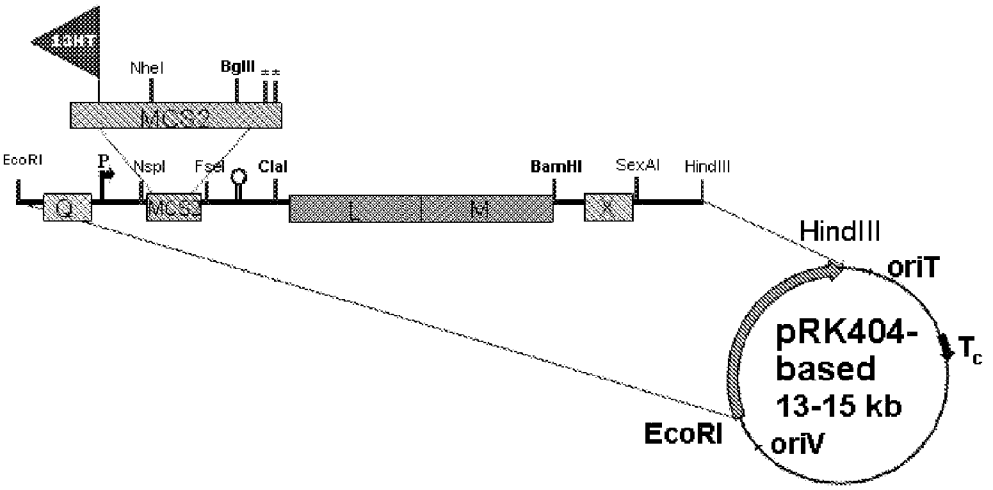
10HT
NheI
BglII
MCS2
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS1
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 6

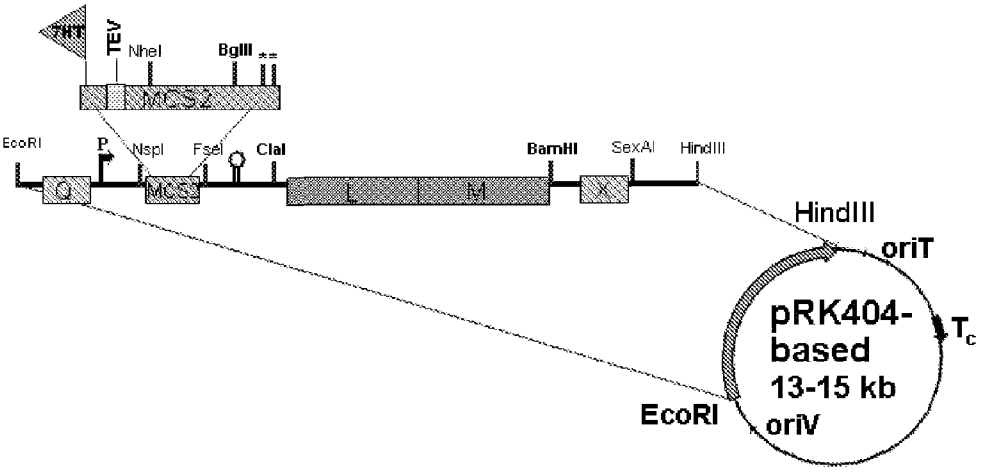
ZHT
TEV
NheI
BglII
MCS2
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
MCS1
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
T_c

FIG. 7

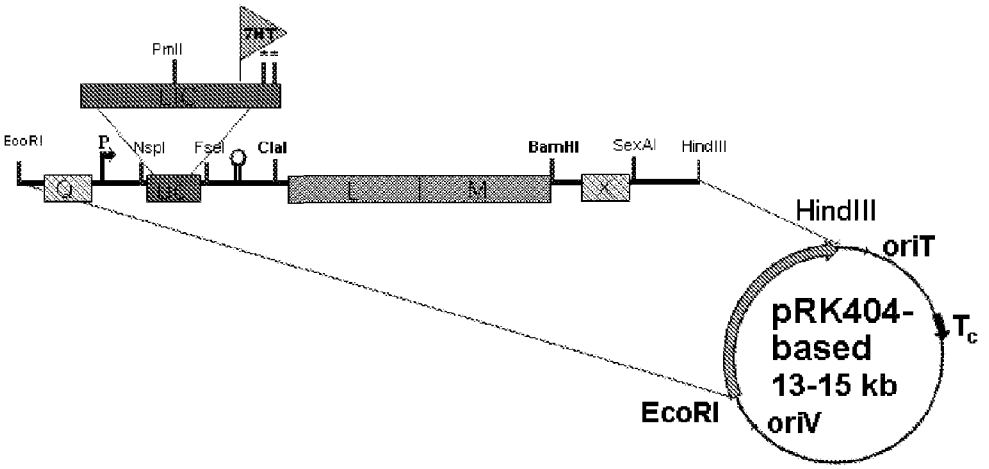
PmlI
ZHT
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
T_c

FIG. 8

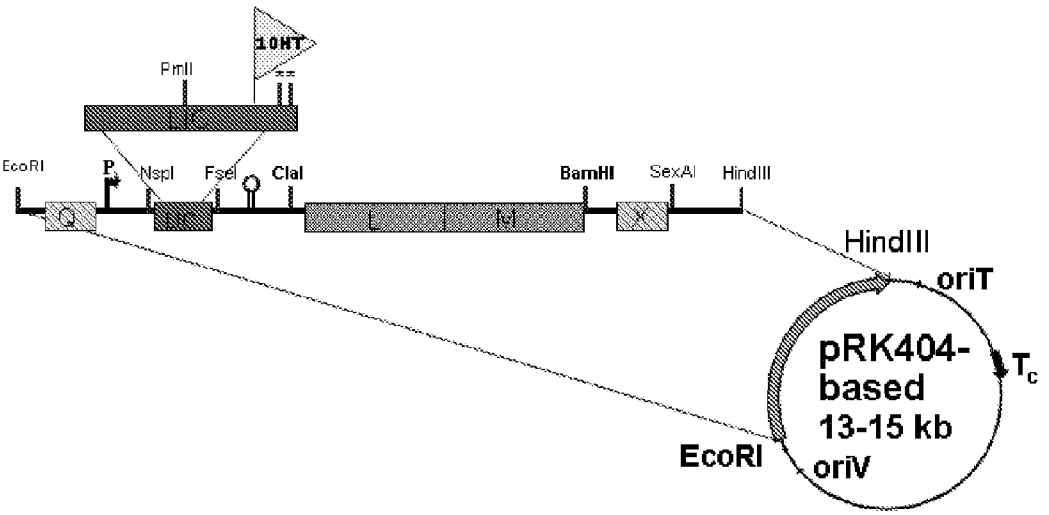
10HT
PmlI
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
N
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
T_c

FIG. 9

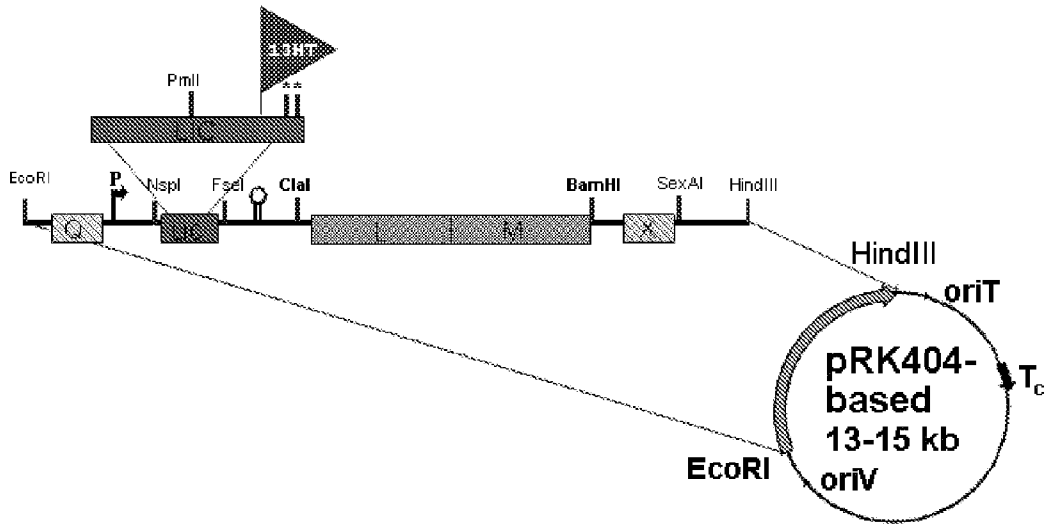
13HT
PmlI
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
N
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
T_c

FIG. 10

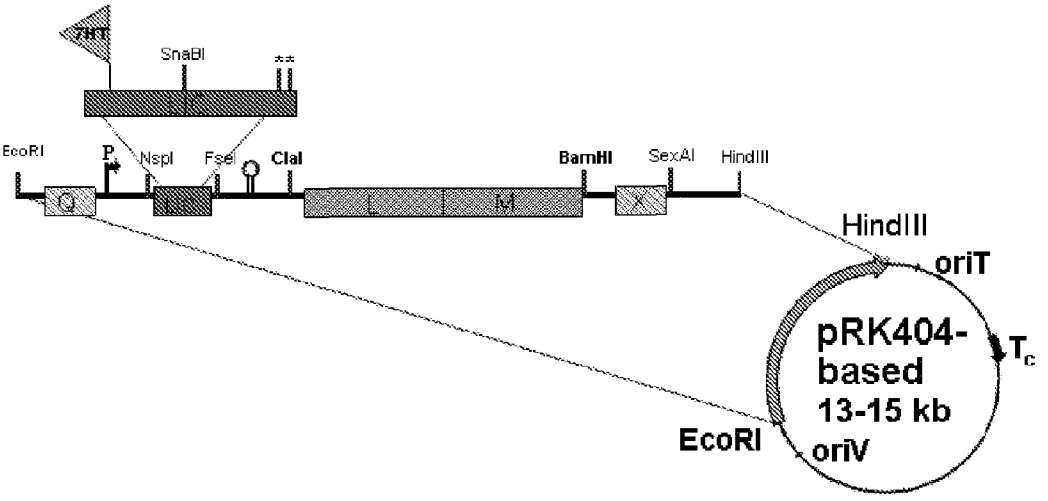
7HT
SnaBI
**
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 11

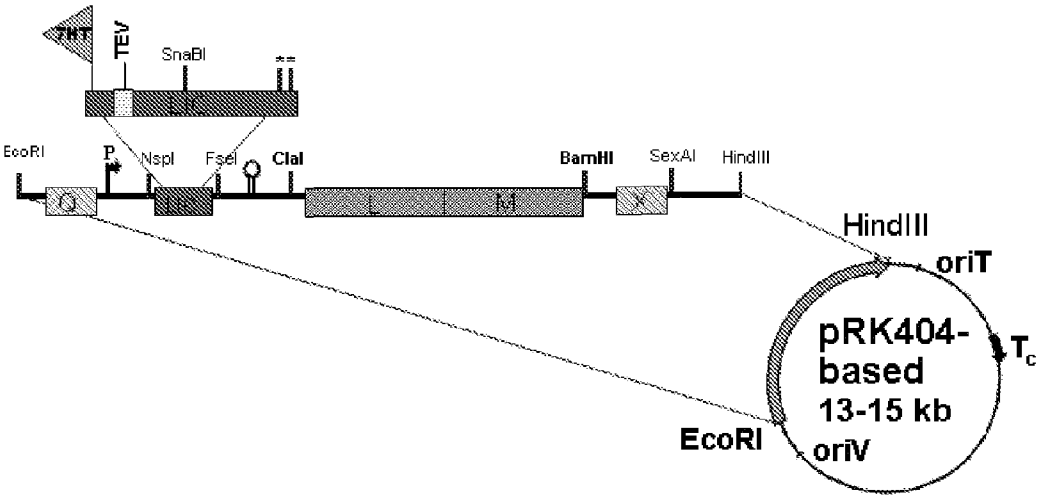
7HT
TEV
SnaBI
**
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
L
M
X
HindIII
oriT
pRK404-
based
13-15 kb
oriV
EcoRI
Tc

FIG. 12

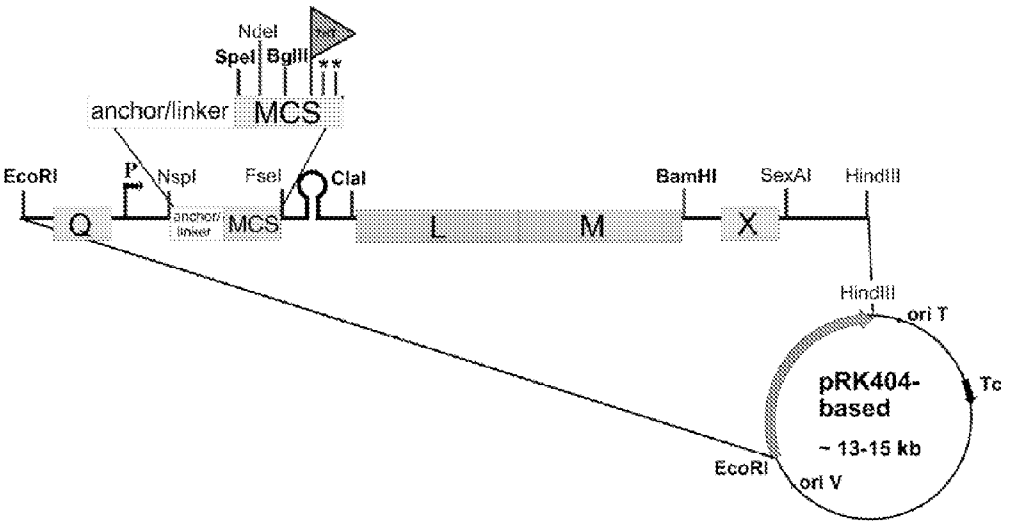
NdeI
SpeI
BglII
anchor/linker
MCS
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
L
M
X
ori T
pRK404-based
~ 13-15 kb
Tc
ori V

FIG. 13

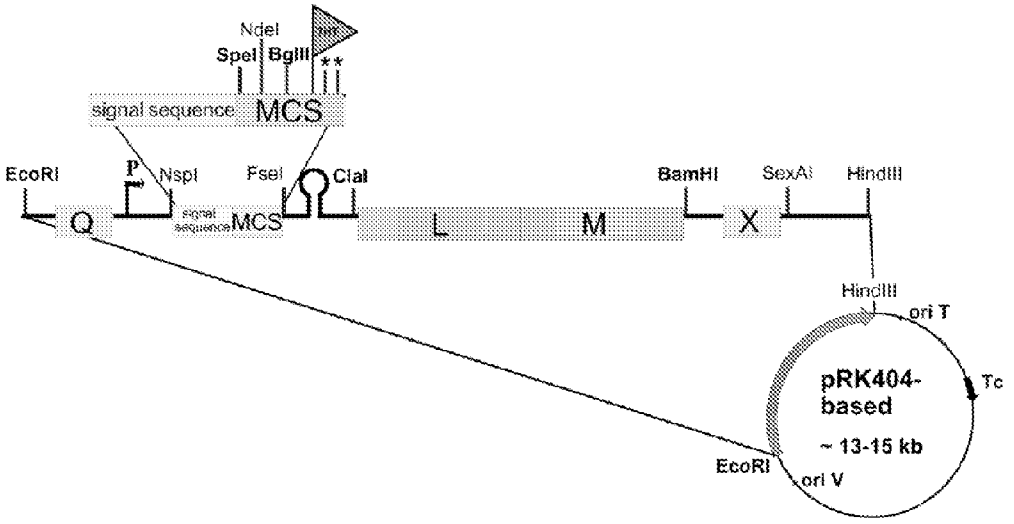
NdeI
SpeI
BglII
signal sequence
MCS
EcoRI
P
NspI
FseI
ClaI
BamHI
SexAI
HindIII
Q
L
M
X
ori T
pRK404-based
~ 13-15 kb
Tc
ori V

FIG. 14

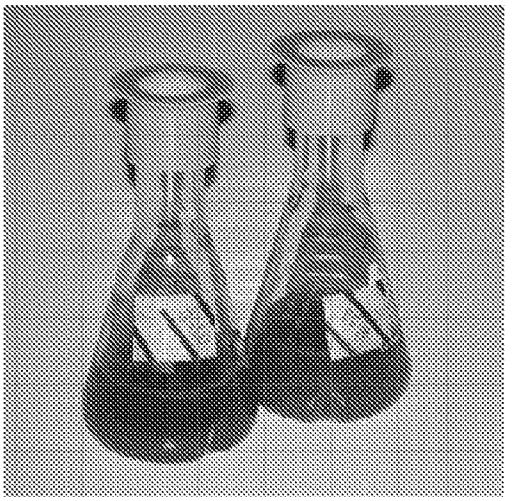

FIG. 15

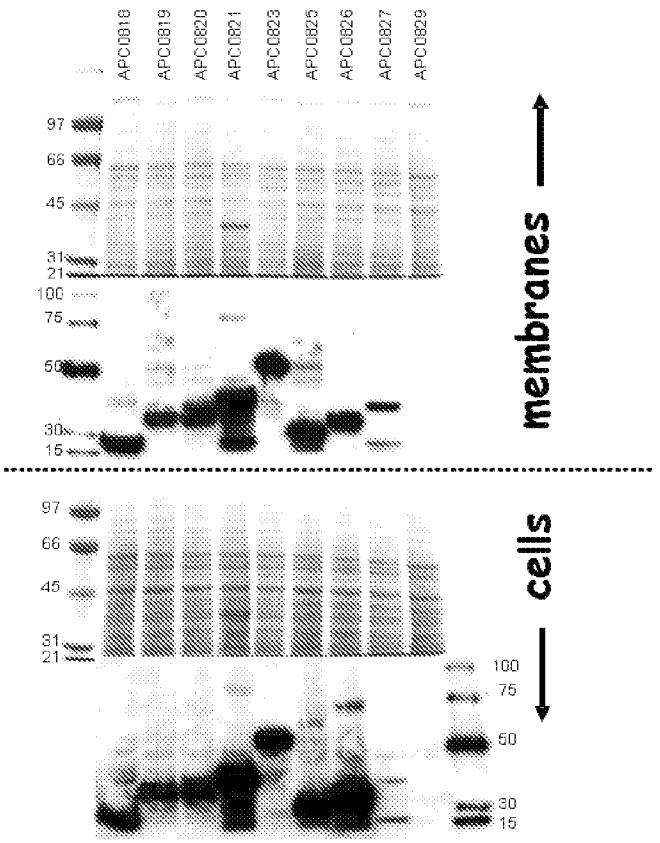
APC0818
APC0819
APC0820
APC0821
APC0823
APC0825
APC0826
APC0827
APC0829
97
66
45
31
21
100
75
50
30
15
membranes
cells

FIG. 16

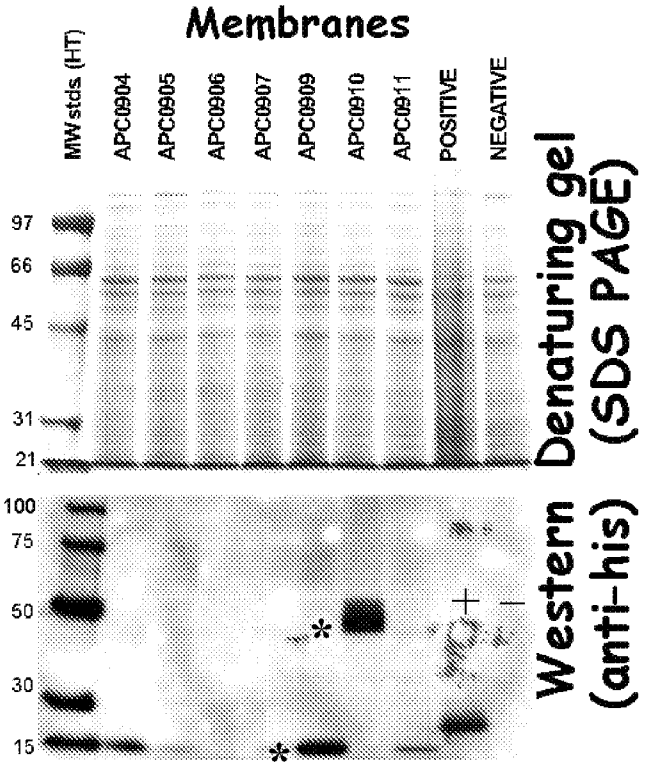
Membranes
MW stds (HT)
APC0904
APC0905
APC0906
APC0907
APC0909
APC0910
APC0911
POSITIVE
NEGATIVE
97
66
45
31
21
100
75
50
30
15
Denaturing gel (SDS PAGE)
Western (anti-his)
+
−

FIG. 17

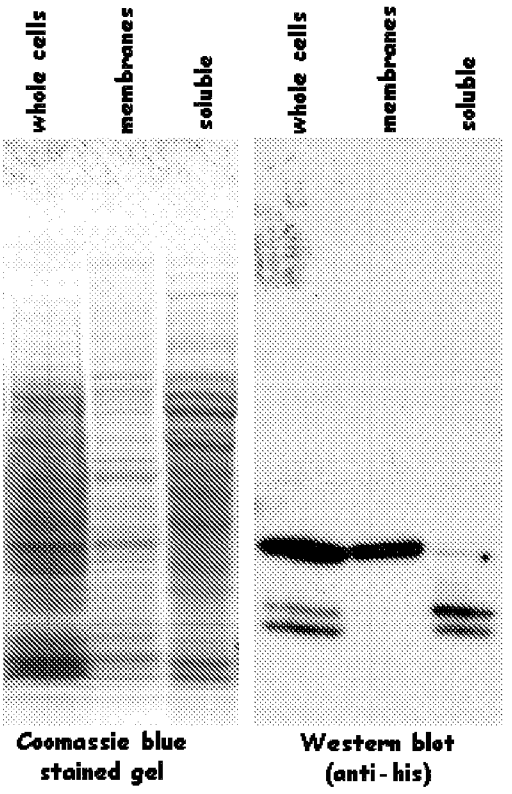
whole cells
membranes
soluble
whole cells
membranes
soluble
Coomassie blue stained gel
Western blot (anti-his)

FIG. 18

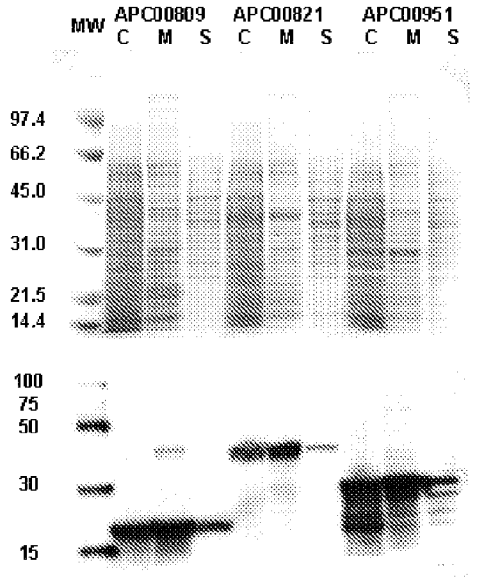
MW
APC00809
APC00821
APC00951
C M S C M S C M S
97.4
66.2
45.0
31.0
21.5
14.4
100
75
50
30
15

FIG. 19

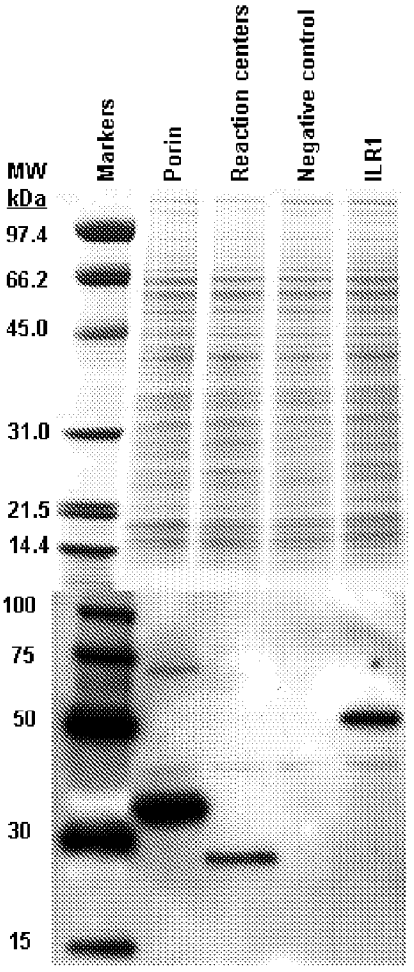
MW
kDa
Markers
Porin
Reaction centers
Negative control
ILR1
97.4
66.2
45.0
31.0
21.5
14.4
100
75
50
30
15

FIG. 20

FIG. 21

FIG. 22

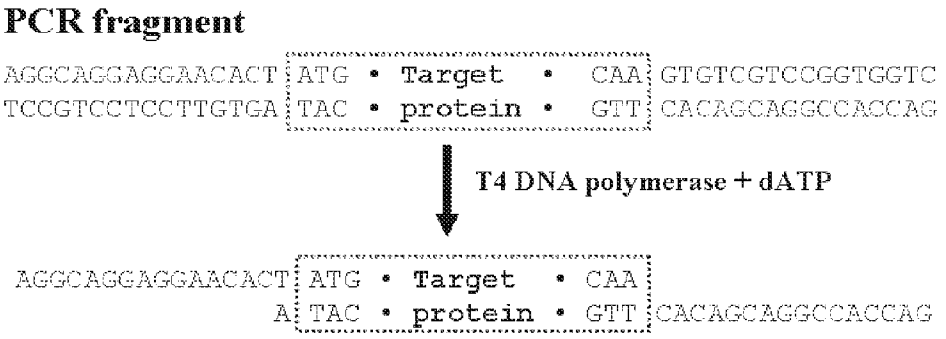

FIG. 23

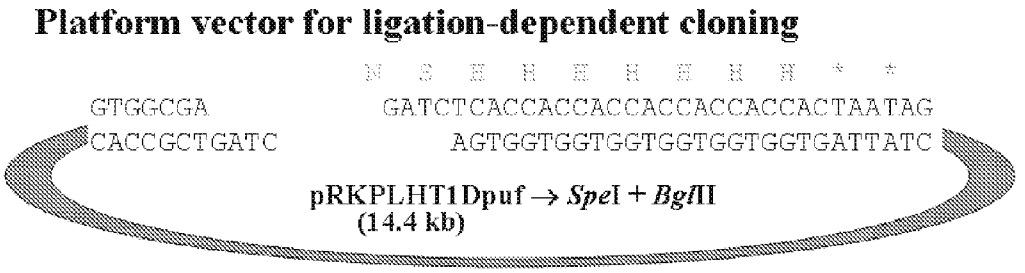
Platform vector for ligation-dependent cloning
N S E H E H E H H * *
GTGGCGA
GATCTCACCACCACCACCACCACTAATAG
CACCGCTGATC
AGTGGTGGTGGTGGTGGTGGTGATTATC
pRKPLHT1Dpuf → SpeI + BglII
(14.4 kb)

FIG. 24

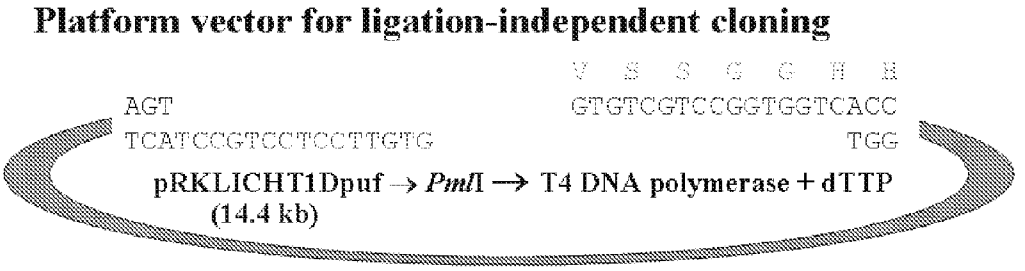
Platform vector for ligation-independent cloning
V S S G G H H
AGT
GTGTCGTCCGGTGGTCACC
TCATCCGTCCTCCTTGTG
TGG
pRKLICHT1Dpuf → PmlI → T4 DNA polymerase + dTTP
(14.4 kb)

FIG. 25

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 *Halobacterium 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 puh and puf 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 (APC00809) 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 (APC00809) 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_ms 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 t-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* ΔΔ11 (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 pRKHTpuf (or a derivative of it) is used to complement, in trans, a strain of *R. sphaeroides* (ΔΔ11) that carries an engineered deletion of the chromosomal copy of this operon. The genes for the LH2 complex are also deleted in strain ΔΔ11, thus the phenotype of this strain is $LH1^{s}$ $LH2^{s}RC^{s}$.

[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 multistep 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 c_y of *R. capsulatus* (13, FIG. **13**). In addition, a vector containing a cleavable, N-terminal signal sequence derived from cytochrome c_2 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 litigation-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 "triaged" 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_1, 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 OD_{600} 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 OD_{600} 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 Electroblotting 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 Anti-polyhistidine 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 μg vial or 15 μL sterile water per 3 μg vial (final concentration 0.2 mg/mL).

[0153] Chemiluminescent detection: Alkali-soluble Casein (Novagen 70955-3; stored at 4EC) 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×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×TBS and 1×TBSTT. They may be prepared by diluting 10×stocks. Filter sterilize the 1×TBSTT.

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

[0157] The Goat Anti-Mouse IgG HRP conjugate 2Eantibody (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×8.5 cm pieces of PVDF that fit purchased mini-gels, use clear plastic 6.5 cm×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×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×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 μL 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×TBSTT to remove unbound 1E antibody.

[0162] Wash for 10 minutes with 15 mL 1×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×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×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 hydropathy plot signatures are selected intitially.

[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

Adapatability 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, *Rhodopseudomonas, 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. sphaeroide*. 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 Rhodobacter 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	Cloning Strategy	Tag Position	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	Cloning Strategy	Tag Position	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	BglII 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)
		BglII 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	BglII restriction site (bold)
tag)	2578–2607	10x Histidine tag (gray)
	2608–2613	Stop codons (italicized)
pRKPLHT13Dpuf	2564–2569	SpeI restriction site (bold)

TABLE 2-continued

Description of cloning sites and affinity tags in the expression vectors

Vector name	DNA sequence position	Site Description
(C-terminal 13 x His tag)	2572–2577	BglII restriction site (bold)
	2578–2616	13x Histidine tag (gray)
	2617–2622	Stop codons (italicized)
pRKMAHT1Dpuf (N-terminal membrane anchor/linker domain; C-terminal 7 x His tag)	2572–2719	Membrane anchor/linker domain
	2720–2725	SpeI restriction site (bold)
	2736–2741	BglII restriction site (bold)
	2742–2762	7 x Histidine tag (gray)
	2763–2768	Stop codons (italicized)
pRKSSHT1Dpuf (N-terminal signal sequence; C-terminal 7 x His tag)	2639–2644	SpeI restriction site (bold)
	2655–2660	BglII restriction site (bold)
	2573–2638	Signal sequence (underlined)
	2661–2681	7 x Histidine tag (gray)
	2682–2687	Stop codons (italicized)

MATERIALS AND METHODS

Vector Construction

[0179] Construction of pRKHTPL1Dpuf: The small EcoRI-HindIII fragment of plasmid pRKHTMHBgl [11] was subcloned into pBluescript SK+(Stratagene, Inc.). The insert of the resulting plasmid was digested with NspI and FseI to excise the pufB and pufA genes encoding subunits of the light-harvesting I antennae complex. That fragment was replaced by a synthetic oligonucleotide cassette:

5′ TGGAGGATCGCCATGCACCACCACCACCACCACCAC**GCTAGC**GCGGGCAGATCTTGA*GGCCGG* 3′
3′ GTACACCTCCTAGCGGTACGTGGTGGTGGTGGTGGTG**CGATCG**CGCCCGTCTAGAACT*CC* 5′

[0180] encoding an NspI site (double underlined), an ATG start codon (dashed underline), a 7×Histidine tag (CAC, gray), an FseI site (italics), and unique NheI (bold) and BglII (underlined) sites for cloning of foreign genes. This modified region was then excised as an EcoRI-ClaI fragment and was swapped for the existing EcoRI-ClaI fragment of pRKPLHT1Dpuf.

[0181] Construction of pRKHTTEVPL1Dpuf: The procedure was the same as that used for the construction of pRKHTPL1Dpuf above, with the exception that the synthetic oligonucleotide cassette used was:

5′ TGGAGGATCGCCATGCACCACCACCACCACCACCACGAGAACCTGTACTTCCAGTCC
3′ GTACACCTCCTAGCGGTACGTGGTGGTGGTGGTGGTGCTCTTGGACATGAAGGTCAGG

GCTAGCTAGGGAAGATCTTGA*GGCCGG* 3′

CGATCGATCCCTTCTAGAACT*CC* 5′

[0182] encoding an NspI site (double underlined), a ribosome binding site (thick underline), an ATG start codon (dashed underline), a 7×Histidine tag (CAC, black), a segment encoding the recognition site for the protease from Tobacco Etch Virus (TEV;

[0183] dotted emphasis) an FseI site (italics), and unique NheI (bold) and BglII (underlined) sites for cloning of foreign genes. This modified region was then excised as an EcoRI-ClaI fragment and was swapped for the existing EcoRI-ClaI fragment of pRKPLHT1Dpuf.

[0184] Construction of pRKPLHT10Dpuf: The small EcoRI-HindIII fragment of plasmid pRKHTMHBgl was subcloned into pBS+(Stratagene, Inc.). The insert of the resulting plasmid was digested with NspI and FseI to excise the pufB and pufA genes encoding subunits of the light-harvesting I antennae complex. That fragment was replaced by a synthetic oligonucleotide cassette:

```
5'-CTAGTTCCA@TGATAGATCTCACCACCACCACCACCACCACCACCA
CCACTAATAGGCCGG-3'

AAGGTATACTATC@GAGTGGTGGTGGTGGTGGTGGTGGTGGTGATT
ATCC-3'
@ indicates text missing or illegible when filed
```

[0185] encoding an NspI site (blue), an ATG start codon (pink), a 7×Histidine tag (CAC, black), an FseI site (orange), and unique NheI (teal) and BglII (purple) sites for cloning of foreign genes. This modified region was then excised as an EcoRI-ClaI fragment and was swapped for the existing EcoRI-ClaI fragment of pRKPLHT1Dpuf.

[0186] Construction of pRKPLHT13Dpuf: The procedure was the same as that used for the construction of pRKPLHT10Dpuf above, with the exception that the synthetic oligonucleotide cassette used contained codons for 13 histidine residues instead of 10.

[0187] Construction of pRKLICHT1Dpuf: The small EcoRI-HindIII fragment of plasmid pRKPLHT1Dpuf was subcloned into pBluescript SK+(Stratagene, Inc.). The insert of the resulting plasmid was digested with PflMI and BglII to excise the multiple cloning site region of this expression vector. The DNA was then treated with mung bean nuclease to generate blunt ends, then the plasmid was treated with T4 DNA polymerase in the presence of dTTP to generate the desired overhangs that were complementary to the following synthetic oligonucleotide cassette:

```
5'-AACCCACGCCACCAGTAGGCAGGAGGAACACGTGTCGTCCGGTGG-
3'

3'-ATCCGTCCTCCTTGTGCACAGCAGGCCACCAG-5'
```

[0188] This cassette encodes a ribosome binding site (magenta) and a PmlI site (dark red), and it was annealed to the modified plasmid. This modified cloning region was then excised as an EcoRI-ClaI fragment and was swapped for the existing EcoRI-ClaI fragment in a version of pRKPLHT1Dpuf in which an existing PmlI site had been repaired. The unique PmlI site located in the cloning site region of the resulting expression vector pRKLICHT1Dpuf facilitates the linearization of the plasmid prior to treatment with T4 polymerase to generate overhangs for ligation-independent cloning.

[0189] Construction of pRKLICHT10Dpuf: This plasmid was constructed in a manner analogous to the construction of pRKLICHT1Dpuf above, with the exception that the initial manipulations were performed on a plasmid that contained the small EcoRI-HindIII fragment of plasmid pRKPLHT10Dpuf subcloned into pBluescript SK+(Stratagene, Inc.).

[0190] Construction of pRKLICHT13Dpuf: This plasmid was constructed in a manner analogous to the construction of pRKLICHT1Dpuf above, with the exception that the initial manipulations were performed on a plasmid that contained the small EcoRI-HindIII fragment of plasmid pRKPLHT13Dpuf subcloned into pBluescript SK+(Stratagene, Inc.)

[0191] Construction of pRKHTLIC1Dpuf: Expression vector pRKHTPL1Dpuf was digested with NheI and BglII to excise the cloning site region of the plasmid. The DNA was then treated with mung bean nuclease to generate blunt ends, then the plasmid was treated with T4 DNA polymerase in the presence of dCTP to generate overhangs that were complementary to a synthetic oligonucleotide cassette:

```
5'-GCCTATTCCAATCCTACGTAGAAGGGAAGATC-3'

3'-GGATAAGGTTAGGATGCATCTTCCCTTCTAGAA-5'
```

[0192] The cassette includes a unique SnaBI site (dark green) that facilitates the linearization of the plasmid prior to treatment with T4 polymerase to generate overhangs for ligation-independent cloning.

[0193] Construction of pRKHTTEVLIC1Dpuf: Expression vector pRKHTPL1Dpuf was digested with NheI and BglII to excise the cloning site region of the plasmid. The DNA was then treated with mung bean nuclease to generate blunt ends, then the plasmid was treated with T4 DNA polymerase in the presence of dCTP to generate overhangs that were complementary to a synthetic oligonucleotide cassette:

```
5'-GAGAACCTGTACTTCCAATCCTTTACGTAGAAATAGGGAAGATC-
3'

3'-TCTTGGACATGAAGGTTAGGAAATGCATCTTTATCCCTTCTAGAA-
5'
```

[0194] The cassette includes a region that encodes a recognition site for the TEV protease and a unique SnaBI site (dark green) that facilitates the linearization of the plasmid prior to treatment with T4 polymerase to generate overhangs for ligation-independent cloning.

Design of Oligonucleotide Primers for Gene Amplification and Cloning.

[0195] The choice of cloning strategy (ligation-dependent or ligation-independent) and specific vector (N- or C-terminal tag; protease site, etc.) will dictate the composition of the oligonucleotides used to amplify the target gene. It is also important to determine whether there are any codons in the first 50 that are extremely rare in *Rhodobacter* (e.g., TTA). If so, one may want to consider cloning a paralog or homolog that lacks rare codons. For a higher-throughput approach to designing sets of oligonucleotides for the cloning of multiple target genes, primer generator tools [10] can be used instead of relying on manual design. Two examples include one for ligation-dependent cloning and the other for ligation-independent cloning, both using vectors with C-terminal, non-cleavable polyhistidine tags—of typical oligonucleotide design for amplification of target genes to be compatible with one of the platform vectors of the *Rhodobacter* system.

Ligation-Dependent Cloning Using pRKPLHT1Dpuf.

[0196] For ligation-dependent cloning, decide which enzymes will be used to insert the gene into the expression vector, choosing those enzymes for which there are no sites in the target gene. The SpeI site in pRKPLHT1Dpuf is also compatible with XbaI and AvrII overhangs. The BglII site in pRKPLHT1Dpuf is compatible with BamHI and BclI. Note that combining the BclI overhang with that of BglII pro-

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_ms 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 (pRKPLICHT1Dpuf) 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+μL 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 μL pRKPLHT1Dpuf

[0206] 2.5 μL SpeI

[0207] 2.5 μL BglII

[0208] 8 μL Promega buffer B

[0209] 42 μL sterile ddH_20

[0210] For a total volume of 80 μL

[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 μg/μL.

[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 μL pRKLICHT1Dpuf (1/5 of the yield of plasmid DNA from a standard miniprep protocol) with PmlI in 70 μL 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 μL 100 mM dTTP

[0223] 2 μL 100 mM DTT

[0224] 4 μL 10×T4 polymerase reaction buffer

[0225] 1 unit T4 DNA polymerase

[0226] Total volume of 40 μL

[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 ± His tag for Ligation Independent Cloning

pRKHTLIC1Dpuf.seq Length: 14386

```
   1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG
```

-continued

```
1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCATG[redacted]GCCT ATTCCAATCC
2601 TACGTAGAAG GGAAGATCTT GAGGGCCGGC CCTCCGTCGC GGGCGGCACC
2651 CACGCCCGCa tCGATTCCAA GGTTCAGCCA TTGAGACGGC TCCGCTTCGC
2701 GCGCAAGCGC GGGTTGGGCC GACTGCAAGC GGAGAGGGAA GCATGGCACT
2751 GCTCAGCTTC GAGCGAAAAT ATCGCGTGCC GGGGGGCACG CTGGTCGGCG
2801 GAAACCTGTT CGACTTCTGG GTCGGCCCTT TCTATGTCGG CTTCTTCGGG
2851 GTTGCGACGT TTTTCTTCGC GGCCCTGGGT ATCATTCTGA TTGCCTGGAG
2901 TGCCGTACTC CAGGGTACCT GGAACCCCCA ACTCATCTCT GTCTACCCGC
2951 CGGCCCTTGA ATATGGCCTG GGAGGTGCAC CCCTCGCAAA AGGCGGGCTG
3001 TGGCAGATCA TCACGATCTG CGCCACTGGT GCCTTCGTCA GCTGGGCGCT
3051 GCGCGAAGTC GAAATCTGCC GTAAGCTGGG CATCGGGTAC CACATCCCGT
3101 TCGCCTTCGC GTTCGCCATC CTGGCCTACC TGACGCTGGT GCTGTTCCGC
3151 CCGGTGATGA TGGGCGCCTG GGGCTATGCC TTCCCCTACG GGATCTGGAC
3201 GCACCTCGAC TGGGTGTCGA ACACGGGCTA CACCTACGGC AACTTCCACT
3251 ACAACCCTGC CCACATGATC GCCATCTCGT TCTTCTTCAC GAACGCGCTG
3301 GCTCTGGCGC TGCACGGCGC CCTTGTGCTC TCCGCGGCCA ACCCCGAGAA
3351 GGGCAAGGAA ATGCGGACGC CGGATCACGA GGATACGTTC TTCCGCGATC

3401 TGGTCGGCTA CTCGATCGGG ACGCTCGGCA TCCACCGCCT CGGCCTGCTG
3451 CTCTCGCTGA GCGCCGTCTT CTTCAGCGCC CTCTGCATGA TCATTACCGG
3501 CACCATCTGG TTCGATCAGT GGGTCGACTG GTGGCAATGG TGGGTGAAGC
3551 TGCCGTGGTG GGCGAACATC CCGGGAGGCA TCAATGGCTG AGTATCAGAA
3601 CATCTTCTCC CAGGTCCAGG TCCGCGGACC GGCCGACCTG GGGATGACCG
3651 AAGACGTCAA CCTGGCCAAC CGTTCGGGCG TCGGTCCCTT CTCGACCCTG
3701 CTCGGCTGGT TCGGCAACGC CCAGCTCGGC CCGATCTATC TCGGCTCGCT
3751 CGGCGTCCTG TCCCTCTTCT CGGGCCTGAT GTGGTTCTTC ACCATCGGGA
3801 TCTGGTTCTG GTATCAGGCG GGCTGGAACC CGGCCGTCTT CCTGCGCGAC
3851 CTGTTCTTCT TCTCGCTCGA GCCGCCGGCA CCCGAATACG GTCTGTCCTT
3901 CGCGGCTCCG CTGAAGGAAG GCGGGCTGTG GCTGATCGCG TCGTTCTTCA
3951 TGTTCGTCGC GGTCTGGTCC TGGTGGGGCC GCACCTATCT CCGCGCTCAG
4001 GCGCTGGGCA TGGGCAAGCA CACCGCCTGG GCGTTCCTCT CGGCCATCTG
4051 GCTGTGGATG GTGCTGGGCT TCATCCGTCC GATCCTCATG GGGTCCTGGT
4101 CGGAAGCGGT TCCCTACGGC ATCTTCTCGC ACCTCGACTG GACGAACAAC
4151 TTCTCGCTCG TCCACGGCAA CCTGTTCTAC AACCCCTTCC ACGGTCTCTC
4201 GATCGCCTTC CTCTACGGGT CGGCCCTGCT CTTCGCGATG CACGGTGCGA
4251 CCATCCTCGC GGTCTCCCGC TTCGGCGGCG AGCGCGAGCT GGAGCAGATC
4301 GCCGACCGCG GGACGGCAGC GGAGCGGGCC GCCCTCTTCT GGCGCTGGAC
4351 CATGGGTTTC AACGCCACGA TGGAAGGCAT CCACCGCTGG GCCATCTGGA
4401 TGGCGGTCCT CGTGACCCTC ACCGGCGGCA TCGGgATCCT GCTCTCGGGC
4451 ACGGTCGTGG ACAACTGGTA CGTCTGGGGC CAGAACCACG GCATGGCGCC
4501 GCTGAACTGA GGAGCGATCA CAATGGCTGA CAAGACCATC TTCAACGATC
4551 ACCTCAACAC CAATCCGAAG ACCAACCTTC GCCTCTGGGT CGCTTTCCAG
4601 ATGATGAAGG GTGCGGGCTG GGCTGGCGGC GTGTTCTTCG GGACGCTCCT
4651 TCTCATCGGG TTCTTCCGGG TGGTCGGGCG GATGCTTCCG ATCCAGGAGA
4701 ACCAGGCTCC GGCGCCGAAC ATCACCGGCG CTCTGGAGAC CGGGATCGAG
4751 CTGATCAAGC ATCTCGTCTG AGACAAGTCT CGGGGCAGGG CGGCGCGAGG
4801 CCGCCCGCTC CTCCAAGTCC GGGCCATATC GCCGGCCCGG GTCCGGGGCG
4851 ACACCACAGC CCGGTTCCCT TCCTGTTGGC GACAGGGACC TGGTGCCGTG
4901 TGGAAGACCG CACGGCACCC TTTTGACATT CACGGGAGGC TCTGATGACC
4951 AATCCCACCC CGCGACCCGA AACCCCGCTT TTGGATCGCG TCTGCTGCCC
5001 GGCCGACATG AAGGCGCTGA GTGACGCCGA ACTGGAGCGG CTGGCCGACG
5051 AAGTGCGTTC CGAGGTCAGT GATAGGGGTA GTTTCTTATT TTAGGCAGTT
```

-continued

```
5101 TATATGAAAT TAAGACATGC AGATGTCACA GTGGATATTG AACTGGTCTC
5151 GAAAGCTCAA TATCCCCCAA AGCACAAGCA CAAACTTCGA CATCATGCAG
5201 AAGCGTTTCC CGAAccgcgt cttcgacgtg ggcatcgccg agcagcatgc
5251 cgtgaccttc gcggccggcc tcgCcgGGGc cggGatgaag cccttctgcg
5301 cgatctattc ctcgttcctg caacggggtt acgaccagat cgcccatgac
5351 gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct
5401 cgtgggggcc gatggcgcga cccatgcggg ggccTTCGAC GTTGGCTTCA
5451 TCACTTCGCT GCCCAACATG ACCGTGATGG CCGCGGCCGA CGAGGCCGAG
5501 CTCATCCACA TGATCgcCAC CGCCGTGGCC TTCGACGAGG GCCCCATCGC
5551 CTTcCGCTTC CCGCGGGGCG AGGGGGTGGG CGTCGAGATG CCCGAGCGCG
5601 GGACGGTGCT GGAGCCCGgC CGGGGCCGCG TGGTGCGCGA AGGGACggat
5651 gtcgcgatcc tctccttCGG CGCGcATCTG CAcGAGgccT TGcAGGCggc
5701 GAAACTTCTC GAGGccGAGG GGGTGAGCGT GACCGTGGCC GACgcccgCT
5751 TctCgCgCCC GCTCgAcACG GGGCTCATCG ACCAGCTCGT gcGCCATCAC
5801 GCGGCGCTGG TAACGGTGGA GCATGGGGCc ATGGGCGGCT TCGGCGCCCA
5851 TGTCATGCAC TATCTCGCCA ATTCCGGCGG CTTCGACGGG GGCCTCGCGC
5901 TCCGGGTCAT GACGCTGCCC GACCGCTTCA TCGAGCAGGC GAGCCCCGAG
5951 gACATGTATG CCGATGCGGG GCTGCGGGCC GAGGATATCA AGCTTGGCGT
6001 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT
6051 CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA
6101 ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC
6151 AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG
6201 GGGAGAGGCG GTTTGCGTAT TGGGCGCTCG GTCTTGCCTT GCTCGTCGGT
6251 GATGTACTTC ACCAGCTCCG CGAAGTCGCT CTTCTTGATG GAGCGCATGG
6301 GGACGTGCTT GGCAATCACG CGCACCCCCC GGCCGTTTTA GCGGCTAAAA
6351 AAGTCATGGC TCTGCCCTCG GGCGGACCAC GCCCATCATG ACCTTGCCAA
6401 GCTCGTCCTG CTTCTCTTCG ATCTTCGCCA GCAGGGCGAG GATCGTGGCA
6451 TCACCGAACC GCGCCGTGCG CGGGTCGTCG GTGAGCCAGA GTTTCAGCAG
6501 GCCGCCCAGG CGGCCCAGGT CGCCATTGAT GCGGGCCAGC TCGCGGACGT
6551 GCTCATAGTC CACGACGCCC GTGATTTTGT AGCCCTGGCC GACGGCCAGC
6601 AGGTAGGCCG ACAGGCTCAT GCCGGCCGCC GCCGCCTTTT CCTCAATCGC
6651 TCTTCGTTCG TCTGGAAGGC AGTACACCTT GATAGGTGGG CTGCCCTTCC
6701 TGGTTGGCTT GGTTTCATCA GCCATCCGCT TGCCCTCATC TGTTACGCCG
6751 GCGGTAGCCG GCCAGCCTCG CAGAGCAGGA TTCCCGTTGA GCACCGCCAG

6801 GTGCGAATAA GGGACAGTGA AGAAGGAACA CCCGCTCGCG GGTGGGCCTA
6851 CTTCACCTAT CCTGCCCGGC TGACGCCGTT GGATACACCA AGGAAAGTCT
6901 ACACGAACCC TTTGGCAAAA TCCTGTATAT CGTGCGAAAA AGGATGGATA
6951 TACCGAAAAA ATCGCTATAA TGACCCCGAA GCAGGGTTAT GCAGCGGAAA
7001 AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG
7051 CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAAACGCCT
7101 GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA
7151 TTTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA
7201 CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT
7251 TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT
7301 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC
7351 AGTGAGCGAG GAAGCGGAAG AGCGCCAGAA GGCCGCCAGA GAGGCCGAGC
7401 GCGGCCGTGA GGCTTGGACG CTAGGGCAGG GCATGAAAAA GCCCGTAGCG
7451 GGCTGCTACG GGCGTCTGAC GCGGTGGAAA GGGGGAGGGG ATGTTGTCTA
7501 CATGGCTCTG CTGTAGTGAG TGGGTTGCGC TCCGGCAGCG GTCCTGATCA
7551 ATCGTCACCC TTTCTCGGTC CTTCAACGTT CCTGACAACG AGCCTCCTTT
7601 TCGCCAATCC ATCGACAATC ACCGCGAGTC CCTGCTCGAA CGCTGCGTCC
7651 GGACCGGCTT CGTCGAAGGC GTCTATCGCG GCCCGCAACA GCGGCGAGAG
7701 CGGAGCCTGT TCAACGGTGC CGCCGCGCTC GCCGGCATCG CTGTCGCCGG
7751 CCTGCTCCTC AAGCACGGCC CCAACAGTGA AGTAGCTGAT TGTCATCAGC
7801 GCATTGACGG CGTCCCCGGC CGAAAAACCC GCCTCGCAGA GGAAGCGAAG
7851 CTGCGCGTCG GCCGTTTCCA TCTGCGGTGC GCCCGGTCGC GTGCCGGCAT
7901 GGATGCGCGC GCCATCGCGG TAGGCGAGCA GCGCCTGCCT GAAGCTGCGG
7951 GCATTCCCGA TCAGAAATGA GCGCCAGTCG TCGTCGGCTC TCGGCACCGA
8001 ATGCGTATGA TTCTCCGCCA GCATGGCTTC GGCCAGTGCG TCGAGCAGCG
8051 CCCGCTTGTT CCTGAAGTGC CAGTAAAGCG CCGGCTGCTG AACCCCCAAC
8101 CGTTCCGCCA GTTTGCGTGT CGTCAGACCG TCTACGCCGA CCTCGTTCAA
8151 CAGGTCCAGG GCGGCACGGA TCACTGTATT CGGCTGCAAC TTTGTCATGC
8201 TTGACACTTT ATCACTGATA AACATAATAT GTCCACCAAC TTATCAGTGA
8251 TAAAGAATCC GCGCGTTCAA TCGGACCAGC GGAGGCTGGT CCGGAGGCCA
8301 GACaTGAAAC CCAACATACC CCTGATCGTA ATTCTGAGCA CTGTCGCGCT
8351 CGACGCTGTC GGCATCGGCC TGATTATGCC GGTGCTGCCG GGCCTCCTGC
8401 GCGATCTGGT TCACTCGAAC GACGTCACCG CCCACTATGG CATTCTGCTG
8451 GCGCTGTATG CGTTGGTGCA ATTTGCCTGC GCACCTGTGC TGGGCGCGCT
```

-continued

```
 8501 GTCGGATCGT TTCGGGCGGC GGCCAATCTT GCTCGTCTCG CTGGCCGGCG
 8551 CCACTGTCGA CTACGCCATC ATGGCGACAG CGCCTTTCCT TTGGGTTCTC
 8601 TATATCGGGC GGATCGTGGC CGGCATCACC GGGGCGACTG GGGCGGTAGC
 8651 CGGCGCTTAT ATTGCCGATA TCACTGATGG CGATGAGCGC GCGCGGCACT
 8701 TCGGCTTCAT GAGCGCCTGT TTCGGGTTCG GGATGGTCGC GGGACCTGTG
 8751 CTCGGTGGGC TGATGGGCGG TTTCTCCCCC CACGCTCCGT TCTTCGCCGC
 8801 GGCAGCCTTG AACGGCCTCA ATTTCCTGAC GGGCTGTTTC CTTTTGCCGG
 8851 AGTCGCACAA AGGCGAACGC CGGCCGTTAC GCCGGGAGGC TCTCAACCCG
 8901 CTCGCTTCGT TCCGGTGGGC CCGGGGCATG ACCGTCGTCG CCGCCCTGAT
 8951 GGCGGTCTTC TTCATCATGC AACTTGTCGG ACAGGTGCCG GCCGCGCTTT
 9001 GGGTCATTTT CGGCGAGGAT CGCTTTCACT GGGACGCGAC CACGATCGGC
 9051 ATTTCGCTTG CCGCATTTGG CATTCTGCAT TCACTCGCCC AGGCAATGAT
 9101 CACCGGCCCT GTAGCCGCCC GGCTCGGCGA AAGGCGGGCA CTCATGCTCG
 9151 GAATGATTGC CGACGGCACA GGCTACATCC TGCTTGCCTT CGCGACACGG
 9201 GGATGGATGG CGTTCCCGAT CATGGTCCTG CTTGCTTCGG GTGGCATCGG
 9251 AATGCCGGCG CTGCAAGCAA TGTTGTCCAG GCAGGTGGAT GAGGAACGTC
 9301 AGGGGCAGCT GCAAGGCTCA CTGGCGGCGC TCACCAGCCT GACCTCGATC
 9351 GTCGGACCCC TCCTCTTCAC GGCGATCTAT GCGGCTTCTA TAACAACGTG
 9401 GAACGGGTGG GCATGGATTG CAGGCGCTGC CCTCTACTTG CTCTGCCTGC
 9451 CGGCGCTGCG TCGCGGGCTT TGGAGCGGCG CAGGGCAACG AGCCGATCGC
 9501 TGATCGTGGA AACGATAGGC CTATGCCATG CGGGTCAAGG CGACTTCCGG
 9551 CAAGCTATAC GCGCCCTAGG AGTGCGGTTG GAACGTTGGC CCAGCCAGAT
 9601 ACTCCCGATC ACGAGCAGGA CGCCGATGAT TTGAAGCGCA CTCAGCGTCT
 9651 GATCCAAGAA CAACCATCCT AGCAACACGG CGGTCCCCGG GCTGAGAAAG
 9701 CCCAGTAAGG AAACAACTGT AGGTTCGAGT CGCGAGATCC CCCGGAACCA
 9751 AAGGAAGTAG GTTAAACCCG CTCCGATCAG GCCGAGCCAC GCCAGGCCGA
 9801 GAACATTGGT TCCTGTAGGC ATCGGGATTG GCGGATCAAA CACTAAAGCT
 9851 ACTGGAACGA GCAGAAGTCC TCCGGCCGCC AGTTGCCAGG CGGTAAAGGT
 9901 GAGCAGAGGC ACGGGAGGTT GCCACTTGCG GGTCAGCACG GTTCCGAACG
 9951 CCATGGAAAC CGCCCCCGCC AGGCCCGCTG CGACGCCGAC AGGATCTAGC
10001 GCTGCGTTTG GTGTCAACAC CAACAGCGCC ACGCCCGCAG TTCCGCAAAT
10051 AGCCCCCAGG ACCGCCATCA ATCGTATCGG GCTACCTAGC AGAGCGGCAG
10101 AGATGAACAC GACCATCAGC GGCTGCACAG CGCCTACCGT CGCCGCGACC
10151 CCGCCCGGCA GGCGGTAGAC CGAAATAAAC AACAAGCTCC AGAATAGCGA

10201 AATATTAAGT GCGCCGAGGA TGAAGATGCG CATCCACCAG ATTCCCGTTG
10251 GAATCTGTCG GACGATCATC ACGAGCAATA AACCCGCCGG CAACGCCCGC
10301 AGCAGCATAC CGGCGACCCC TCGGCCTCGC TGTTCGGGCT CCACGAAAAC
10351 GCCGGACAGA TGCGCCTTGT GAGCGTCCTT GGGGCCGTCC TCCTGTTTGA
10401 AGACCGACAG CCCAATGATC TCGCCGTCGA TGTAGGCGCC GAATGCCACG
10451 GCATCTCGCA ACCGTTCAGC GAACGCCTCC ATGGGCTTTT TCTCCTCGTG
10501 CTCGTAAACG GACCCGAACA TCTCTGGAGC TTTCTTCAGG GCCGACAATC
10551 GGATCTCGCG GAAATCCTGC ACGTCGGCCG CTCCAAGCCG TCGAATCTGA
10601 GCCTTAATCA CAATTGTCAA TTTTAATCCT CTGTTTATCG GCAGTTCGTA
10651 GAGCGCGCCG TGCGTCCCGA GCGATACTGA GCGAAGCAAG TGCGTCGAGC
10701 AGTGCCCGCT TGTTCCTGAA ATGCCAGTAA AGCGCTGGCT GCTGAACCCC
10751 CAGCCGGAAC TGACCCCACA AGGCCCTAGC GTTTGCAATG CACCAGGTCA
10801 TCATTGACCC AGGCGTGTTC CACCAGGCCG CTGCCTCGCA ACTCTTCGCA
10851 GGCTTCGCCG ACCTGCTCGC GCCACTTCTT CACGCGGGTG GAATCCGATC
10901 CGCACATGAG GCGGAAGGTT TCCAGCTTGA GCGGGTACGG CTCCCGGTGC
10951 GAGCTGAAAT AGTCGAACAT CCGTCGGGCC GTCGGCGACA GCTTGCGGTA
11001 CTTCTCCCAT ATGAATTTCG TGTAGTGGTC GCCAGCAAAC AGCACGACGA
11051 TTTCCTCGTC GATCAGGACC TGGCAACGGG ACGTTTTCTT GCCACGGTCC
11101 AGGACGCGGA AGCGGTGCAG CAGCGACACC GATTCCAGGT GCCCAACGCG
11151 GTCGGACGTG AAGCCCATCG CCGTCGCCTG TAGGCGCGAC AGGCATTCCT
11201 CGGCCTTCGT GTAATACCGG CCATTGATCG ACCAGCCCAG GTCCTGGCAA
11251 AGCTCGTAGA ACGTGAAGGT GATCGGCTCG CCGATAGGGG TGCGCTTCGC
11301 GTACTCCAAC ACCTGCTGCC ACACCAGTTC GTCATCGTCG GCCCGCAGCT
11351 CGACGCCGGT GTAGGTGATC TTCACGTCCT TGTTGACGTG GAAAATGACC
11401 TTGTTTTGCA GCGCCTCGCG CGGGATTTTC TTGTTGCGCG TGGTGAACAG
11451 GGCAGAGCGG GCCGTGTCGT TTGGCATCGC TCGCATCGTG TCCGGCCACG
11501 GCGCAATATC GAACAAGGAA AGCTGCATTT CCTTGATCTG CTGCTTCGTG
11551 TGTTTCAGCA ACGCGGCCTG CTTGGCCTCG CTGACCTGTT TTGCCAGGTC
11601 CTCGCCGGCG GTTTTTCGCT TCTTGGTCGT CATAGTTCCT CGCGTGTCGA
11651 TGGTCATCGA CTTCGCCAAA CCTGCCGCCT CCTGTTCGAG ACGACGCGAA
11701 CGCTCCACGG CGGCCGATGG CGCGGGCAGG GCAGGGGGAG CCAGTTGCAC
11751 GCTGTCGCGC TCGATCTTGG CCGTAGCTTG CTGGACCATC GAGCCGACGG
11801 ACTGGAAGGT TTCGCGGGGC GCACGCATGA CGGTGCGGCT TGCGATGGTT
11851 TCGGCATCCT CGGCGGAAAA CCCCGCGTCG ATCAGTTCTT GCCTGTATGC
```

-continued

```
11901 CTTCCGGTCA AACGTCCGAT TCATTCACCC TCCTTGCGGG ATTGCCCCGA
11951 CTCACGCCGG GGCAATGTGC CCTTATTCCT GATTTGACCC GCCTGGTGCC
12001 TTGGTGTCCA GATAATCCAC CTTATCGGCA ATGAAGTCGG TCCCGTAGAC
12051 CGTCTGGCCG TCCTTCTCGT ACTTGGTATT CCGAATCTTG CCCTGCACGA
12101 ATACCAGCGA CCCCTTGCCC AAATACTTGC CGTGGGCCTC GGCCTGAGAG
12151 CCAAAACACT TGATGCGGAA GAAGTCGGTG CGCTCCTGCT TGTCGCCGGT
12201 CGTGGCCGCG CCAACCTTTG CGATCCGCAA GCGCGCGGTC GCCATCTTCA
12251 CGCTGGAACA GTACGTCGAG GCGGGCATCA TGACCCGCGA GCAATACGAG
12301 GTCATTAAAA GCGCCGTGAT TGATGATATA GCGGCCCGGC TGCTCCTGGT
12351 TCTCGCGCAC CGAAATGGGT GACTTCACCC CGCGCTCTTT GATCGTGGCA
12401 CCGATTTCCG CGATGCTCTC CGGGGAAAAG CCGGGGTTGT CGGCCGTCCG
12451 CGGCTGATGC GGATCTTCGT CGATCAGGTC CAGGTCCAGC TCGATAGGGC
12501 CGGAACCGCC CTGAGACGCC GCAGGAGCGT CCAGGAGGCT CGACAGGTCG
12551 CCGATGCTAT CCAACCCCAG GCCGGACGGC TGCGCCGCGC CTGCGGCTTC
12601 CTGAGCGGCC GCAGCGGTGT TTTTCTTGGT GGTCTTGGCT TGAGCCGCAG
12651 TCATGGGGAA ATCTCCATCT TCGTGAACAC GTAATCAGCC AGGGCGCGAA
12701 CCTCTTTCGA TGCCTTGCGC GCGGCCGTTT TCTTGATCTT CCAGACCGGC
12751 ACACCGGATG CGAGGGCATC GGCGATGCTG CTGCGCAGGC CAACGGTGGC
12801 CGGAATCATC ATCTTGGGGT ACGCGGCCAG CAGCTCGGCT TGGTGGCGCG
12851 CGTGGCGCGG ATTCCGCGCA TCGACCTTGC TGGGCACCAT GCCAAGGAAT
12901 TGCAGCTTGG CGTTCTTCTG GCGCACGTTC GCAATGGTCG TGACCATCTT
12951 CTTGATGCCC TGGATGCTGT ACGCCTCAAG CTCGATGGGG GACAGCACAT
13001 AGTCGGCCGC GAAGAGGGCG GCCGCCAGGC CGACGCCAAG GGTCGGGGCC
13051 GTGTCGATCA GGCACACGTC GAAGCCTTGG TTCGCCAGGG CCTTGATGTT
13101 CGCCCCGAAC AGCTCGCGGG CGTCGTCCAG CGACAGCCGT TCGGCGTTCG
13151 CCAGTACCGG GTTGGACTCG ATGAGGGCGA GGCGCGCGGC CTGGCCGTCG
13201 CCGGCTGCGG GTGCGGTTTC GGTCCAGCCG CCGGCAGGGA CAGCGCCGAA
13251 CAGCTTGCTT GCATGCAGGC CGGTAGCAAA GTCCTTGAGC GTGTAGGACG
13301 CATTGCCCTG GGGGTCCAGG TCGATCACGG CAACCCGCAA GCCGCGCTCG
13351 AAAAAGTCGA AGGCAAGATG CACAAGGGTC GAAGTCTTGC CGACGCCGCC
13401 TTTCTGGTTG GCCGTGACCA AAGTTTTCAT CGTTTGGTTT CCTGTTTTTT
13451 CTTGGCGTCC GCTTCCCACT TCCGGACGAT GTACGCCTGA TGTTCCGGCA
13501 GAACCGCCGT TACCCGCGCG TACCCCTCGG GCAAGTTCTT GTCCTCGAAC
13551 GCGGCCCACA CGCGATGCAC CGCTTGCGAC ACTGCGCCCC TGGTCAGTCC

13601 CAGCGACGTT GCGAACGTCG CCTGTGGCTT CCCATCGACT AAGACGCCCC
13651 GCGCTATCTC GATGGTCTGC TGCCCCACTT CCAGCCCCTG GATCGCCTCC
13701 TGGAACTGGC TTTCGGTAAG CCGTTTCTTC ATGGATAACA CCCATAATTT
13751 GCTCCGCGCC TTGGTTGAAC ATAGCGGTGA CAGCCGCCAG CACATGAGAG
13801 AAGTTTAGCT AAACATTTCT CGCACGTCAA CACCTTTAGC TTCGTCTCTT
13851 CGTCCTTGGC GTAACAAAAC AAAAGCCCGG AAADDGGGCT TTCGTCTCTT
13901 GCCGCTTATG GCTCTGCACC CGGCTCCATC ACCAACAGGT CGCGCACGCG
13951 CTTCACTCGG TTGCGGATCG ACACTGCCAG CCCAACAAAG CCGGTTGCCG
14001 CCGCCGCCAG GATCGCGCCG ATGATGCCGG CCACACCGGC CATCGCCCAC
14051 CAGGTCGCCG CCTTCCGGTT CCATTCCTGC TGGTACTGCT TCGCAATGCT
14101 GGACCTCGGC TCACCATAGG CTGACCGCTC GATGGCGTAT GCCGCTTCTC
14151 CCCTTGGCGT AAAACCCAGC GCCGCAGGCG GCATTGCCAT GCTGCCCGCC
14201 GCTTTCCCGA CCACGACGCG CGCACCAGGC TTGCGGTCCA GACCTTCGGC
14251 CACGGCGAGC TGCGCAAGGA CATAATCAGC CGCCGACTTG GCTCCACGCG
14301 CCTCGATCAG CTCTTGCACT CGCGCGAAAT CCTTGGCCTC CACGGCCGCC
14351 ATGAATCGCG CACGCGGCGA AGGCTCCGCA GGGCCG
```

-continued

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

pRKHTPL1Dpuf.seq Length: 14370

```
   1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA AAGTGTCAAC GTCCGCCCCT
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCAATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCATGGACCA CCACCACCAC CACCAGGCTA GCGCGGGCAG
2601 ATCTTGAGGC CGGCCCTCCG TCGCGGGCGG CACCCACGCC CGCatCGATT
2651 CCAAGGTTCA GCCATTGAGA CGGCTCCGCT TCGCGCGCAA GCGCGGGTTG
2701 GGCCGACTGC AAGCGGAGAG GGAAGCATGG CACTGCTCAG CTTCGAGCGA
2751 AAATATCGCG TGCCGGGGGG CACGCTGGTC GGCGGAAACC TGTTCGACTT
2801 CTGGGTCGGC CCTTTCTATG TCGGCTTCTT CGGGGTTGCG ACGTTTTTCT
2851 TCGCGGCCCT GGGTATCATT CTGATTGCCT GGAGTGCCGT ACTCCAGGGT
2901 ACCTGGAACC CCCAACTCAT CTCTGTCTAC CCGCCGGCCC TTGAATATGG
2951 CCTGGGAGGT GCACCCCTCG CAAAAGGCGG GCTGTGGCAG ATCATCACGA
3001 TCTGCGCCAC TGGTGCCTTC GTCAGCTGGG CGCTGCGCGA AGTCGAAATC
3051 TGCCGTAAGC TGGGCATCGG GTACCACATC CCGTTCGCCT TCGCGTTCGC
3101 CATCCTGGCC TACCTGACGC TGGTGCTGTT CCGCCCGGTG ATGATGGGCG
3151 CCTGGGGCTA TGCCTTCCCC TACGGGATCT GGACGCACCT CGACTGGGTG
3201 TCGAACACGG GCTACACCTA CGGCAACTTC CACTACAACC CTGCCCACAT
3251 GATCGCCATC TCGTTCTTCT TCACGAACGC GCTGGCTCTG GCGCTGCACG
3301 GCGCCCTTGT GCTCTCCGCG GCCAACCCCG AGAAGGGCAA GGAAATGCGG
3351 ACGCCGGATC ACGAGGATAC GTTCTTCCGC GATCTGGTCG GCTACTCGAT
```

-continued

```
3401 CGGGACGCTC GGCATCCACC GCCTCGGCCT GCTGCTCTCG CTGAGCGCCG
3451 TCTTCTTCAG CGCCCTCTGC ATGATCATTA CCGGCACCAT CTGGTTCGAT
3501 CAGTGGGTCG ACTGGTGGCA ATGGTGGGTG AAGCTGCCGT GGTGGGCGAA
3551 CATCCCGGGA GGCATCAATG GCTGAGTATC AGAACATCTT CTCCCAGGTC
3601 CAGGTCCGCG GACCGGCCGA CCTGGGGATG ACCGAAGACG TCAACCTGGC
3651 CAACCGTTCG GGCGTCGGTC CCTTCTCGAC CCTGCTCGGC TGGTTCGGCA
3701 ACGCCCAGCT CGGCCCGATC TATCTCGGCT CGCTCGGCGT CCTGTCCCTC
3751 TTCTCGGGCC TGATGTGGTT CTTCACCATC GGGATCTGGT TCTGGTATCA
3801 GGCGGGCTGG AACCCGGCCG TCTTCCTGCG CGACCTGTTC TTCTTCTCGC
3851 TCGAGCCGCC GGCACCCGAA TACGGTCTGT CCTTCGCGGC TCCGCTGAAG
3901 GAAGGCGGGC TGTGGCTGAT CGCGTCGTTC TTCATGTTCG TCGCGGTCTG
3951 GTCCTGGTGG GGCCGCACCT ATCTCCGCGC TCAGGCGCTG GGCATGGGCA
4001 AGCACACCGC CTGGGCGTTC CTCTCGGCCA TCTGGCTGTG GATGGTGCTG
4051 GGCTTCATCC GTCCGATCCT CATGGGGTCC TGGTCGGAAG CGGTTCCCTA
4101 CGGCATCTTC TCGCACCTCG ACTGGACGAA CAACTTCTCG CTCGTCCACG
4151 GCAACCTGTT CTACAACCCC TTCCACGGTC TCTCGATCGC CTTCCTCTAC
4201 GGGTCGGCCC TGCTCTTCGC GATGCACGGT GCGACCATCC TCGCGGTCTC
4251 CCGCTTCGGC GGCGAGCGCG AGCTGGAGCA GATCGCCGAC CGCGGGACGG
4301 CAGCGGAGCG GGCCGCCCTC TTCTGGCGCT GGACCATGGG TTTCAACGCC
4351 ACGATGGAAG GCATCCACCG CTGGGCCATC TGGATGGCGG TCCTCGTGAC
4401 CCTCACCGGC GGCATCGGgA TCCTGCTCTC GGGCACGGTC GTGGACAACT
4451 GGTACGTCTG GGGCCAGAAC CACGGCATGG CGCCGCTGAA CTGAGGAGCG
4501 ATCACAATGG CTGACAAGAC CATCTTCAAC GATCACCTCA ACACCAATCC
4551 GAAGACCAAC CTTCGCCTCT GGGTCGCTTT CCAGATGATG AAGGGTGCGG
4601 GCTGGGCTGG CGGCGTGTTC TTCGGGACGC TCCTTCTCAT CGGGTTCTTC
4651 CGGGTGGTCG GGCGGATGCT TCCGATCCAG GAGAACCAGG CGCCGGCGCC
4701 GAACATCACC GGCGCTCTGG AGACCGGGAT CGAGCTGATC AAGCATCTCG
4751 TCTGAGACAA GTCTCGGGGC AGGGCGGCGC GAGGCCGCCC GCTCCTCCAA
4801 GTCCGGGCCA TATCGCCGGC CCGGGTCCGG GGCGACACCA CAGCCCGGTT
4851 CCCTTCCTGT TGGCGACAGG GACCTGGTGC CGTGTGGAAG ACCGCACGGC
4901 ACCCTTTTGA CATTCACGGG AGGCTCTGAT GACCAATCCC ACCCCGCGAC
4951 CCGAAACCCC GCTTTTGGAT CGCGTCTGCT GCCCGGCCGA CATGAAGGCG
5001 CTGAGTGACG CCGAACTGGA GCGGCTGGCC GACGAAGTGC GTTCCGAGGT
5051 CAGTGATAGG GGTAGTTTCT TATTTTAGGC AGTTTATATG AAATTAAGAC

5101 ATGCAGATGT CACAGTGGAT ATTGAACTGG TCTCGAAAGC TCAATATCCC
5151 CCAAAGCACA AGCACAAACT TCGACATCAT GCAGAAGCGT TTCCCGAAcc
5201 gcgtcttcga cgtgggcatc gccgagcagc atgccgtgac cttcgcggcc
5251 ggcctcgCcg GGGccggGat gaagcccttc tgcgcgatct attcctcgtt
5301 cctgcaacgg ggttacgacc agatcgccca tgacgtggcg ctgcagaacc
5351 ttcccgtccg cttcgtgatc gaccgggcgg ggctcgtggg ggccgatggc
5401 gcgacccatg cgggggccTT CGACGTTGGC TTCATCACTT CGCTGCCCAA
5451 CATGACCGTG ATGGCCGCGG CCGACGAGGC CGAGCTCATC CACATGATCg
5501 cCACCGCCGT GGCCTTCGAC GAGGGCCCCA TCGCCTTcCG CTTCCCGCGG
5551 GGCGAGGGGG TGGGCGTOGA GATGCCCGAG CGCGGGACGG TGCTGGAGCC
5601 CGgCCGGGGC CGCGTGGTGC GCGAAGGGAC ggatgtcgcg atcctctcct
5651 tCGGCGCGcA TCTGCAcGAG gccTTGcAGG CggcGAAACT TCTCGAGGcc
5701 GAGGGGGTGA GCGTGACCGT GGCCGACgcc cgCTTctCgC gCCCGCTCgA
5751 cACGGGGCTC ATCGACCAGC TCGTgcGCCA TCACGCGGCG CTGGTAACGG
5801 TGGAGCAGGG GGCcATGGGC GGCTTCGGCG CCCATGTCAT GCACTATCTC
5851 GCCAATTCCG GCGGCTTCGA CGGGGGCCTC GCGCTCCGGG TCATGACGCT
5901 GCCCGACCGC TTCATCGAGC AGGCGAGCCC CGAGgACATG TATGCCGATG
5951 CGGGGCTGCG GGCCGAGGAT ATCAAGCTTG GCGTAATCAT GGTCATAGCT
6001 GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC AACATACGAG
6051 CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT GAGCTAACTC
6101 ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC
6151 GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC
6201 GTATTGGGCG CTCGGTCTTG CCTTGCTCGT CGGTGATGTA CTTCACCAGC
6251 TCCGCGAAGT CGCTCTTCTT GATGGAGCGC ATGGGGACGT GCTTGGCAAT
6301 CACGCGCACC CCCCGGCCGT TTTAGCGGCT AAAAAAGTCA TGGCTCTGCC
6351 CTCGGGCGGA CCACGCCCAT CATGACCTTG CCAAGCTCGT CCTGCTTCTC
6401 TTCGATCTTC GCCAGCAGGG CGAGGATCGT GGCATCACCG AACCGCGCCG
6451 TGCGCGGGTC GTCGGTGAGC CAGAGTTTCA GCAGGCCGCC CAGGCGGCCC
6501 AGGTCGCCAT TGATGCGGGC CAGCTCGCGG ACGTGCTCAT AGTCCACGAC
6551 GCCCGTGATT TTGTAGCCCT GGCCGACGGC CAGCAGGTAG GCCGACAGGC
6601 TCATGCCGGC CGCCGCCGCC TTTTCCTCAA TCGCTCTTCG TTCGTCTGGA
6651 AGGCAGTACA CCTTGATAGG TGGGCTGCCC TTCCTGGTTG GCTTGGTTTC
6701 ATCAGCCATC CGCTTGCCCT CATCTGTTAC GCCGGCGGTA GCCGGCCAGC
6751 CTCGCAGAGC AGGATTCCCG TTGAGCACCG CCAGGTGCGA ATAAGGGACA
```

-continued

```
 6801 GTGAAGAAGG AACACCCGCT CGCGGGTGGG CCTACTTCAC CTATCCTGCC
 6851 CGGCTGACGC CGTTGGATAC ACCAAGGAAA GTCTACACGA ACCCTTTGGC
 6901 AAAATCCTGT ATATCGTGCG AAAAAGGATG GATATACCGA AAAAATCGCT
 6951 ATAATGACCC CGAAGCAGGG TTATGCAGCG GAAAAGCGCC ACGCTTCCCG
 7001 AAGGGAGAAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT CGGAACAGGA
 7051 GAGCGCACGA GGGAGCTTCC AGGGGGAAAC GCCTGGTATC TTTATAGTCC
 7101 TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG TGATGCTCGT
 7151 CAGGGGGGCG GAGCCTATGG AAAAACGCCA GCAACGCGGC CTTTTTACGG
 7201 TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC CTGCGTTATC
 7251 CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG
 7301 CTCGCCGCAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG
 7351 GAAGAGCGCC AGAAGGCCGC CAGAGAGGCC GAGCGCGGCC GTGAGGCTTG
 7401 GACGCTAGGG CAGGGCATGA AAAAGCCCGT AGCGGGCTGC TACGGGCGTC
 7451 TGACGCGGTG GAAAGGGGGA GGGGATGTTG TCTACATGGC TCTGCTGTAG
 7501 TGAGTGGGTT GCGCTCCGGC AGCGGTCCTG ATCAATCGTC ACCCTTTCTC
 7551 GGTCCTTCAA CGTTCCTGAC AACGAGCCTC CTTTTCGCCA ATCCATCGAC
 7601 AATCACCGCG AGTCCCTGCT CGAACGCTGC GTCCGGACCG GCTTCGTCGA
 7651 AGGCGTCTAT CGCGGCCCGC AACAGCGGCG AGAGCGGAGC CTGTTCAACG
 7701 GTGCCGCCGC GCTCGCCGGC ATCGCTGTCG CCGGCCTGCT CCTCAAGCAC
 7751 GGCCCCAACA GTGAAGTAGC TGATTGTCAT CAGCGCATTG ACGGCGTCCC
 7801 CGGCCGAAAA ACCCGCCTCG CAGAGGAAGC GAAGCTGCGC GTCGGCCGTT
 7851 TCCATCTGCG GTGCGCCCGG TCGCGTGCCG GCATGGATGC GCGCGCCATC
 7901 GCGGTAGGCG AGCAGCGCCT GCCTGAAGCT GCGGGCATTC CCGATCAGAA
 7951 ATGAGCGCCA GTCGTCGTCG GCTCTCGGCA CCGAATGCGT ATGATTCTCC
 8001 GCCAGCATGG CTTCGGCCAG TGCGTCGAGC AGCGCCCGCT TGTTCCTGAA
 8051 GTGCCAGTAA AGCGCCGGCT GCTGAAACCC CAACCGTTCC GCCAGTTTGC
 8101 GTGTCGTCAG ACCGTCTACG CCGACCTCGT TCAACAGGTC CAGGGCGGCA
 8151 CGGATCACTG TATTCGGCTG CAACTTTGTC ATGCTTGACA CTTTATCACT
 8201 GATAAACATA ATATGTCCAC CAACTTATCA GTGATAAAGA ATCCGCGCGT
 8251 TCAATCGGAC CAGCGGAGGC TGGTCCGGAG GCCAGACaTG AAACCCAACA
 8301 TACCCCTGAT CGTAATTCTG AGCACTGTCG CGCTCGACGC TGTCGGCATC
 8351 GGCCTGATTA TGCCGGTGCT GCCGGGCCTC CTGCGCGATC TGGTTCACTC
 8401 GAACGACGTC ACCGCCCACT ATGGCATTCT GCTGGCGCTG TATGCGTTGG
 8451 TGCAATTTGC CTGCGCACCT GTGCTGGGCG CGCTGTCGGA TCGTTTCGGG

 8501 CGGCGGCCAA TCTTGCTCGT CTCGCTGGCC GGCGCCACTG TCGACTACGC
 8551 CATCATGGCG ACAGCGCCTT TCCTTTGGGT TCTCTATATC GGGCGGATCG
 8601 TGGCCGGCAT CACCGGGGCG ACTGGGGCGG TAGCCGGCGC TTATATTGCC
 8651 GATATCACTG ATGGCGATGA GCGCGCGCGG CACTTCGGCT TCATGAGCGC
 8701 CTGTTTCGGG TTCGGGATGG TCGCGGGACC TGTGCTCGGT GGGCTGATGG
 8751 GCGGTTTCTC CCCCCACGCT CCGTTCTTCG CCGCGGCAGC CTTGAACGGC
 8801 CTCAATTTCC TGACGGGCTG TTTCCTTTTG CCGGAGTCGC ACAAAGGCGA
 8851 ACGCCGGCCG TTACGCCGGG AGGCTCTCAA CCCGCTCGCT TCGTTCCGGT
 8901 GGGCCCGGGG CATGACCGTC GTCGCCGCCC TGATGGCGGT CTTCTTCATC
 8951 ATGCAACTTG TCGGACAGGT GCCGGCCGCG CTTTGGGTCA TTTTCGGCGA
 9001 GGATCGCTTT CACTGGGACG CGACCACGAT CGGCATTTCG CTTGCCGCAT
 9051 TTGGCATTCT GCATTCACTC GCCCAGGCAA TGATCACCGG CCCTGTAGCC
 9101 GCCCGGCTCG GCGAAAGGCG GGCACTCATG CTCGGAATGA TTGCCGACGG
 9151 CACAGGCTAC ATCCTGCTTG CCTTCGCGAC ACGGGGATGG ATGGCGTTCC
 9201 CGATCATGGT CCTGCTTGCT TCGGGTGGCA TCGGAATGCC GGCGCTGCAA
 9251 GCAATGTTGT CCAGGCAGGT GGATGAGGAA CGTCAGGGGC AGCTGCAAGG
 9301 CTCACTGGCG GCGCTCACCA GCCTGACCTC GATCGTCGGA CCCCTCCTCT
 9351 TCACGGCGAT CTATGCGGCT TCTATAACAA CGTGGAACGG GTGGGCATGG
 9401 ATTGCAGGCG CTGCCCTCTA CTTGCTCTGC CTGCCGGCGC TGCGTCGCGG
 9451 GCTTTGGAGC GGCGCAGGGC AACGAGCCGA TCGCTGATCG TGGAAACGAT
 9501 AGGCCTATGC CATGCGGGTC AAGGCGACTT CCGGCAAGCT ATACGCGCCC
 9551 TAGGAGTGCG GTTGGAACGT TGGCCCAGCC AGATACTCCC GATCACGAGC
 9601 AGGACGCCGA TGATTTGAAG CGCACTCAGC GTCTGATCCA AGAACAACCA
 9651 TCCTAGCAAC ACGGCGGTCC CCGGGCTGAG AAAGCCCAGT AAGGAAACAA
 9701 CTGTAGGTTC GAGTCGCGAG ATCCCCCGGA ACCAAAGGAA GTAGGTTAAA
 9751 CCCGCTCCGA TCAGGCCGAG CCACGCCAGG CCGAGAACAT TGGTTCCTGT
 9801 AGGCATCGGG ATTGGCGGAT CAAACACTAA AGCTACTGGA ACGAGCAGAA
 9851 GTCCTCCGGC CGCCAGTTGC CAGGCGGTAA AGGTGAGCAG AGGCACGGGA
 9901 GGTTGCCACT TGCGGGTCAG CACGGTTCCG AACGCCATGG AAACCGCCCC
 9951 CGCCAGGCCC GCTGCGACGC CGACAGGATC TAGCGCTGCG TTTGGTGTCA
10001 ACACCAACAG CGCCACGCCC GCAGTTCCGC AAATAGCCCC CAGGACCGCC
10051 ATCAATCGTA TCGGGCTACC TAGCAGAGCG GCAGAGATGA ACACGACCAT
10101 CAGCGGCTGC ACAGCGCCTA CCGTCGCCGC GACCCCGCCC GGCAGGCGGT
10151 AGACCGAAAT AAACAACAAG CTCCAGAATA GCGAAATATT AAGTGCGCCG
```

-continued

```
10201 AGGATGAAGA TGCGCATCCA CCAGATTCCC GTTGGAATCT GTCGGACGAT
10251 CATCACGAGC AATAAACCCG CCGGCAACGC CCGCAGCAGC ATACCGGCGA
10301 CCCCTCGGCC TCGCTGTTCG GGCTCCACGA AAACGCCGGA CAGATGCGCC
10351 TTGTGAGCGT CCTTGGGGCC GTCCTCCTGT TTGAAGACCG ACAGCCCAAT
10401 GATCTCGCCG TCGATGTAGG CGCCGAATGC CACGGCATCT CGCAACCGTT
10451 CAGCGAACGC CTCCATGGGC TTTTTCTCCT CGTGCTCGTA AACGGACCCG
10501 AACATCTCTG GAGCTTTCTT CAGGGCCGAC AATCGGATCT CGCGGAAATC
10551 CTGCACGTCG GCCGCTCCAA GCCGTCGAAT CTGAGCCTTA ATCACAATTG
10601 TCAATTTTAA TCCTCTGTTT ATCGGCAGTT CGTAGAGCGC GCCGTGCGTC
10651 CCGAGCGATA CTGAGCGAAG CAAGTGCGTC GAGCAGTGCC CGCTTGTTCC
10701 TGAAATGCCA GTAAAGCGCT GGCTGCTGAA CCCCCAGCCG GAACTGACCC
10751 CACAAGGCCC TAGCGTTTGC AATGCACCAG GTCATCATTG ACCCAGGCGT
10801 GTTCCACCAG GCCGCTGCCT CGCAACTCTT CGCAGGCTTC GCCGACCTGC
10851 TCGCGCCACT TCTTCACGCG GGTGGAATCC GATCCGCACA TGAGGCGGAA
10901 GGTTTCCAGC TTGAGCGGGT ACGGCTCCCG GTGCGAGCTG AAATAGTCGA
10951 ACATCCGTCG GGCCGTCGGC GACAGCTTGC GGTACTTCTC CCATATGAAT
11001 TTCGTGTAGT GGTCGCCAGC AAACAGCACG ACGATTTCCT CGTCGATCAG
11051 GACCTGGCAA CGGGACGTTT TCTTGCCACG GTCCAGGACG CGGAAGCGGT
11101 GCAGCAGCGA CACCGATTCC AGGTGCCCAA CGCGGTCGGA CGTGAAGCCC
11151 ATCGCCGTCG CCTGTAGGCG CGACAGGCAT TCCTCGGCCT TCGTGTAATA
11201 CCGGCCATTG ATCGACCAGC CCAGGTCCTG GCAAAGCTCG TAGAACGTGA
11251 AGGTGATCGG CTCGCCGATA GGGGTGCGCT TCGCGTACTC CAACACCTGC
11301 TGCCACACCA GTTCGTCATC GTCGGCCCGC AGCTCGACGC CGGTGTAGGT
11351 GATCTTCACG TCCTTGTTGA CGTGGAAAAT GACCTTGTTT TGCAGCGCCT
11401 CGCGCGGGAT TTTCTTGTTG CGCGTGGTGA ACAGGGCAGA GCGGGCCGTG
11451 TCGTTTGGCA TCGCTCGCAT CGTGTCCGGC CACGGCGCAA TATCGAACAA
11501 GGAAAGCTGC ATTTCCTTGA TCTGCTGCTT CGTGTGTTTC AGCAACGCGG
11551 CCTGCTTGGC CTCGCTGACC TGTTTTGCCA GGTCCTCGCC GGCGGTTTTT
11601 CGCTTCTTGG TCGTCATAGT TCCTCGCGTG TCGATGGTCA TCGACTTCGC
11651 CAAACCTGCC GCCTCCTGTT CGAGACGACG CGAACGCTCC ACGGCGGCCG
11701 ATGGCGCGGG CAGGGCAGGG GGAGCCAGTT GCACGCTGTC GCGCTCGATC
11751 TTGGCCGTAG CTTGCTGGAC CATCGAGCCG ACGGACTGGA AGGTTTCGCG
11801 GGGCGCACGC ATGACGGTGC GGCTTGCGAT GGTTTCGGCA TCCTCGGCGG
11851 AAAACCCCGC GTCGATCAGT TCTTGCCTGT ATGCCTTCCG GTCAAACGTC

11901 CGATTCATTC ACCCTCCTTG CGGGATTGCC CCGACTCACG CCGGGGCAAT
11951 GTGCCCTTAT TCCTGATTTG ACCCGCCTGG TGCCTTGGTG TCCAGATAAT
12001 CCACCTTATC GGCAATGAAG TCGGTCCCGT AGACCGTCTG GCCGTCCTTC
12051 TCGTACTTGG TATTCCGAAT CTTGCCCTGC ACGAATACCA GCGACCCCTT
12101 GCCCAAATAC TTGCCGTGGG CCTCGGCCTG AGAGCCAAAA CACTTGATGC
12151 GGAAGAAGTC GGTGCGCTCC TGCTTGTCGC CGGTCGTGGC CGCGCCAACC
12201 TTTGCGATCC GCAAGCGCGC GGTCGCCATC TTCACGCTGG AACAGTACGT
12251 CGAGGCGGGC ATCATGACCC GCGAGCAATA CGAGGTCATT AAAAGCGCCG
12301 TGATTGATGA TATAGCGGCC CGGCTGCTCC TGGTTCTCGC GCACCGAAAT
12351 GGGTGACTTC ACCCCGCGCT CTTTGATCGT GGCACCGATT TCCGCGATGC
12401 TCTCCGGGGA AAAGCCGGGG TTGTCGGCCG TCCGCGGCTG ATGCGGATCT
12451 TCGTCGATCA GGTCCAGGTC CAGCTCGATA GGGCCGGAAC CGCCCTGAGA
12501 CGCCGCAGGA GCGTCCAGGA GGCTCGACAG GTCGCCGATG CTATCCAACC
12551 CCAGGCCGGA CGGCTGCGCC GCGCCTGCGG CTTCCTGAGC GGCCGCAGCG
12601 GTGTTTTTCT TGGTGGTCTT GGCTTGAGCC GCAGTCATTG GGAAATCTCC
12651 ATCTTCGTGA ACACGTAATC AGCCAGGGCG CGAACCTCTT TCGATGCCTT
12701 GCGCGCGGCC GTTTTCTTGA TCTTCCAGAC CGGCACACCG GATGCGAGGG
12751 CATCGGCGAT GCTGCTGCGC AGGCCAACGG TGGCCGGAAT CATCATCTTG
12801 GGGTACGCGG CCAGCAGCTC GGCTTGGTGG CGCGCGTGGC GCGGATTCCG
12851 CGCATCGACC TTGCTGGGCA CCATGCCAAG GAATTGCAGC TTGGCGTTCT
12901 TCTGGCGCAC GTTCGCAATG GTCGTGACCA TCTTCTTGAT GCCCTGGATG
12951 CTGTACGCCT CAAGCTCGAT GGGGGACAGC ACATAGTCGG CCGCGAAGAG
13001 GGCGGCCGCC AGGCCGACGC CAAGGGTCGG GGCCGTGTCG ATCAGGCACA
13051 CGTCGAAGCC TTGGTTCGCC AGGGCCTTGA TGTTCGCCCC GAACAGCTCG
13101 CGGGCGTCGT CCAGCGACAG CCGTTCGGCG TTCGCCAGTA CCGGGTTGGA
13151 CTCGATGAGG GCGAGGCGCG CGGCCTGGCC GTCGCCGGCT GCGGGTGCGG
13201 TTTCGGTCCA GCCGCCGGCA GGGACAGCGC CGAACAGCTT GCTTGCATGC
13251 AGGCCGGTAG CAAAGTCCTT GAGCGTGTAG GACGCATTGC CCTGGGGGTC
13301 CAGGTCGATC ACGGCAACCC GCAAGCCGCG CTCGAAAAAG TCGAAGGCAA
13351 GATGCACAAG GGTCGAAGTC TTGCCGACGC CGCCTTTCTG GTTGGCCGTG
13401 ACCAAAGTTT TCATCGTTTG GTTTCCTGTT TTTTCTTGGC GTCCGCTTCC
13451 CACTTCCGGA CGATGTACGC CTGATGTTCC GGCAGAACCG CCGTTACCCG
13501 CGCGTACCCC TCGGGCAAGT TCTTGTCCTC GAACGCGGCC CACACGCGAT
13551 GCACCGCTTG CGACACTGCG CCCCTGGTCA GTCCCAGCGA CGTTGCGAAC

13601 GTCGCCTGTG GCTTCCCATC GACTAAGACG CCCCGCGCTA TCTCGATGGT
13651 CTGCTGCCCC ACTTCCAGCC CCTGGATCGC CTCCTGGAAC TGGCTTTCGG
13701 TAAGCCGTTT CTTCATGGAT AACACCCATA ATTTGCTCCG CGCCTTGGTT
13751 GAACATAGCG GTGACAGCCG CCAGCACATG AGAGAAGTTT AGCTAAACAT
13801 TTCTCGCACG TCAACACCTT TAGCCGCTAA AACTCGTCCT TGGCGTAACA
13851 AAACAAAAGC CCGGAAACCG GGCTTTCGTC TCTTGCCSCT TATGGCTCTG
13901 CACCCGGCTC CATCACCAAC AGGTCGCGCA CGCGCTTCAC TCGGTTGCGG
13951 ATCGACACTG CCAGCCCAAC AAAGCCGGTT GCCGCCGCCG CCAGGATCGC
14001 GCCGATGATG CCGGCCACAC CGGCCATCGC CCACCAGGTC GCCGCCTTCC
14051 GGTTCCATTC CTGCTGGTAC TGCTTCGCAA TGCTGGACCT CGGCTCACCA
14101 TAGGCTGACC GCTCGATGGC GTATGCCGCT TCTCCCCTTG GCGTAAAACC
14151 CAGCGCCGCA GGCGGCATTG CCATGCTGCC CGCCGCTTTC CCGACCACGA
14201 CGCGCGCACC AGGCTTGCGG TCCAGACCTT CGGCCACGGC GAGCTGCGCA
14251 AGGACATAAT CAGCCGCCGA CTTGGCTCCA CGCGCCTCGA TCAGCTCTTG
14301 CACTCGCGCG AAATCCTTGG CCTCCACGGC CGCCATGAAT CGCGCACGCG
14351 GCGAAGGCTC CGCAGGGCCG
```

-continued

Broad-host-range expression vector for ligation-independent cloning featuring N-terminal 7 ± His tag followed by cleavage site for Tobacco Etch Virus protease

pRKHTLICTEV1Dpuf.seq Length: 14398

```
   1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TCTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TCGGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCATG[CACCA CCACCACCA CCACCACGAGA ACCTGTACTT
2601 CCAATCCTTT ACGTAGAAAT AGGGAAGATC TTGAGGGCCG GCCCTCCGTC
2651 GCGGGCGGCA CCCACGCCCG CatCGATTCC AAGGTTCAGC CATTGAGACG
2701 GCTCCGCTTC GCGCGCAAGC GCGGGTTGGG CCGACTGCAA GCGGAGAGGG
2751 AAGCATGGCA CTGCTCAGCT TCGAGCGAAA ATATCGCGTG CCGGGGGGCA
2801 CGCTGGTCGG CGGAAACCTG TTCGACTTCT GGGTCGGCCC TTTCTATGTC
2851 GGCTTCTTCG GGGTTGCGAC GTTTTTCTTC GCGGCCCTGG GTATCATTCT
2901 GATTGCCTGG AGTGCCGTAC TCCAGGGTAC CTGGAACCCC CAACTCATCT
2951 CTGTCTACCC GCCGGCCCTT GAATATGGCC TGGGAGGTGC ACCCCTCGCA
3001 AAAGGCGGGC TGTGGCAGAT CATCACGATC TGCGCCACTG GTGCCTTCGT
3051 CAGCTGGGCG CTGCGCGAAG TCGAAATCTG CCGTAAGCTG GGCATCGGGT
3101 ACCACATCCC GTTCGCCTTC GCGTTCGCCA TCCTGGCCTA CCTGACGCTG
3151 GTGCTGTTCC GCCCGGTGAT GATGGGCGCC TGGGGCTATG CCTTCCCCTA
3201 CGGGATCTGG ACGCACCTCG ACTGGGTGTC GAACACGGGC TACACCTACG
3251 GCAACTTCCA CTACAACCCT GCCCACATGA TCGCCATCTC GTTCTTCTTC
3301 ACGAACGCGC TGGCTCTGGC GCTGCACGGC GCCCTTGTGC TCTCCGCGGC
3351 CAACCCCGAG AAGGGCAAGG AAATGCGGAC GCCGGATCAC GAGGATACGT
```

-continued

```
3401 TCTTCCGCGA TCTGGTCGGC TACTCGATCG GGACGCTCGG CATCCACCGC
3451 CTCGGCCTGC TGCTCTCGCT GAGCGCCGTC TTCTTCAGCG CCCTCTGCAT
3501 GATCATTACC GGCACCATCT GGTTCGATCA GTGGGTCGAC TGGTGGCAAT
3551 GGTGGGTGAA GCTGCCGTGG TGGGCGAACA TCCCGGGAGG CATCAATGGC
3601 TGAGTATCAG AACATCTTCT CCCAGGTCCA GGTCCGCGGA CCGGCCGACC
3651 TGGGGATGAC CGAAGACGTC AACCTGGCCA ACCGTTCGGG CGTCGGTCCC
3701 TTCTCGACCC TGCTCGGCTG GTTCGGCAAC GCCCAGCTCG GCCCGATCTA
3751 TCTCGGCTCG CTCGGCGTCC TGTCCCTCTT CTCGGGCCTG ATGTGGTTCT
3801 TCACCATCGG GATCTGGTTC TGGTATCAGG CGGGCTGGAA CCCGGCCGTC
3851 TTCCTGCGCG ACCTGTTCTT CTTCTCGCTC GAGCCGCCGG CACCCGAATA
3901 CGGTCTGTCC TTCGCGGCTC CGCTGAAGGA AGGCGGGCTG TGGCTGATCG
3951 CGTCGTTCTT CATGTTCGTC GCGGTCTGGT CCTGGTGGGG CCGCACCTAT
4001 CTCCGCGCTC AGGCGCTGGG CATGGGCAAG CACACCGCCT GGGCGTTCCT
4051 CTCGGCCATC TGGCTGTGGA TGGTGCTGGG CTTCATCCGT CCGATCCTCA
4101 TGGGGTCCTG 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 CCGCCCTCTT
4351 CTGGCGCTGG ACCATGGGTT TCAACGCCAC GATGGAAGGC ATCCACCGCT
4401 GGGCCATCTG GATGGCGGTC CTCGTGACCC TCACCGGCGG CATCGGgATC
4451 CTGCTCTCGG GCACGGTCGT GGACAACTGG TACGTCTGGG GCCAGAACCA
4501 CGGCATGGCG CCGCTGAACT GAGGAGCGAT CACAATGGCT GACAAGACCA
4551 TCTTCAACGA TCACCTCAAC ACCAATCCGA AGACCAACCT TCGCCTCTGG
4601 GTCGCTTTCC AGATGATGAA GGGTGCGGGC TGGGCTGGCG GCGTGTTCTT
4651 CGGGACGCTC CTTCTCATCG GGTTCTTCCG GGTGGTCGGG CGGATGCTTC
4701 CGATCCAGGA GAACCAGGCT CCGGCGCCGA ACATCACCGG CGCTCTGGAG
4751 ACCGGGATCG AGCTGATCAA GCATCTCGTC TGAGACAAGT CTCGGGGCAG
4801 GGCGGCGCGA GGCCGCCCGC TCCTCCAAGT CCGGGCCATA TCGCCGGCCC
4851 GGGTCCGGGG CGACACCACA GCCCGGTTCC CTTCCTGTTG GCGACAGGGA
4901 CCTGGTGCCG TGTGGAAGAC CGCACGGCAC CCTTTTGACA TTCACGGGAG
4951 GCTCTGATGA CCAATCCCAC CCCGCGACCC GAAACCCCGC TTTTGGATCG
5001 CGTCTGCTGC CCGGCCGACA TGAAGGCGCT GAGTGACSCC GAACTGGAGC
5051 GGCTGGCCGA CGAAGTGCGT TCCGAGGTCA GTGATAGGGG TAGTTTCTTA

5101 TTTTAGGCAG TTTATATGAA ATTAAGACAT GCAGATGTCA CAGTGGATAT
5151 TGAACTGGTC TCGAAAGCTC AATATCCCCC AAAGCACAAG CACAAACTTC
5201 GACATCATGC AGAAGCGTTT CCCGAAccgc gtcttcgacg tgggcatcgc
5251 cgagcagcat gccgtgacct tcgcggccgg cctcgCcgGG GccggGatga
5301 agcccttctg cgcgatctat tcctcgttcc tgcaacgggg ttacgaccag
5351 atcgcccatg acgtggcgct gcagaacctt cccgtccgct tcgtgatcga
5401 ccgggcgggg ctcgtggggg ccgatggcgc gacccatgcg ggggccTTCG
5451 ACTTTGGCTT CATCACTTCG CTGCCCAACA TGACCGTGAT GGCCGCGGCC
5501 GACGAGGCCG AGCTCATCCA CATGATCgcC ACCGCCGTGG CCTTCGACGA
5551 GGGCCCCATC GCCTTcCGCT TCCCGCGGGG CGAGGGGGTG GGCGTCGAGA
5601 TGCCCGAGCG CGGGACGGTG CTGGAGCCCG gCCGGGGCCG CGTGGTGCGC
5651 GAAGGGACgg atgtcgcgat cctctccttC GGCGCGcATC TGCAcGAGgc
5701 cTTGcAGGCg gcGAAACTTC TCGAGGccGA GGGGGTGAGC GTGACCGTGG
5751 CCGACgcccg CTTctCgCgC CCGCTCgAcA CGGGGCTCAT CGACCAGCTC
5801 GTgcGCCATC ACGCGGCGCT GGTAACGGTG GAGCAGGGGG CcATGGGCGG
5851 CTTCGGCGCC CATGTCATGC ACTATCTCGC CAATTCCGGC GGCTTCGACG
5901 GGGGCCTCGC GCTCCGGGTC ATGACGCTGC CCGACCGCTT CATCGAGCAG
5951 GCGAGCCCCG AGgACATGTA TGCCGATGCG GGGCTGCGGG CCGAGGATAT
6001 CAAGCTTGGC GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT
6051 CCGCTCACAA TTCCACACAA CATACGAGCC GGAAGCATAA AGTGTAAAGC
6101 CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG TTGCGCTCAC
6151 TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC
6201 GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGGCGCT CGGTCTTGCC
6251 TTGCTCGTCG GTGATGTACT TCACCAGCTC CGCGAAGTCG CTCTTCTTGA
6301 TGGAGCGCAT GGGGACGTGC TTGGCAATCA CGCGCACCCC CCGGCCGTTT
6351 TAGCGGCTAA AAAAGTCATG GCTCTGCCCT CGGGCGGACC ACGCCCATCA
6401 TGACCTTGCC AAGCTCGTCC TGCTTCTCTT CGATCTTCGC CAGCAGGGCG
6451 AGGATCGTGG CATCACCGAA CCGCGCCGTC CGCGGGTCGT CGGTGAGCCA
6501 GAGTTTCAGC AGGCCGCCCA GGCGGCCCAG GTCGCCATTG ATGCGGGCCA
6551 GCTCGCGGAC GTGCTCATAG TCCACGACGC CCGTGATTTT GTAGCCCTGG
6601 CCGACGGCCA GCAGGTAGGC CGACAGGCTC ATGCCGGCCG CCGCCGCCTT
6651 TTCCTCAATC GCTCTTCGTT CGTCTGGAAG GCAGTACACC TTGATAGGTG
6701 GGCTGCCCTT CCTGGTTGGC TTGGTTTCAT CAGCCATCCG CTTGCCCTCA
6751 TCTGTTACGC CGGCGGTAGC CGGCCAGCCT CGCAGAGCAG GATTCCCGTT
```

-continued

```
 6801 GAGCACCGCC AGGTGCGAAT AAGGGACAGT GAAGAAGGAA CACCCGCTCG
 6851 CGGGTGGGCC TACTTCACCT ATCCTGCCCG GCTGACGCCG TTGGATACAC
 6901 CAAGGAAAGT CTACACGAAC CCTTTGGCAA AATCCTGTAT ATCGTGCGAA
 6951 AAAGGATGGA TATACCGAAA AAATCGCTAT AATGACCCCG AAGCAGGGTT
 7001 ATGCAGCGGA AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA
 7051 TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG
 7101 GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA
 7151 CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA
 7201 AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT
 7251 TTGCTCACAT GTTCTTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT
 7301 ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA
 7351 GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCAG AAGGCCGCCA
 7401 GAGAGGCCGA GCGCGGCCGT GAGGCTTGGA CGCTAGGGCA GGGCATGAAA
 7451 AAGCCCGTAG CGGGCTGCTA CGGGCGTCTG ACGCGGTGGA AAGGGGGAGG
 7501 GGATGTTGTC TACATGGCTC TGCTGTAGTG AGTGGGTTGC GCTCCGGCAG
 7551 CGGTCCTGAT CAATCGTCAC CCTTTCTCGG TCCTTCAACG TTCCTGACAA
 7601 CGAGCCTCCT TTTCGCCAAT CCATCGACAA TCACCGCGAG TCCCTGCTCG
 7651 AACGCTGCGT CCGGACCGGC TTCGTCGAAG GCGTCTATCG CGGCCCGCAA
 7701 CAGCGGCGAG AGCGGAGCCT GTTCAACGGT GCCGCCGCGC TCGCCGGCAT
 7751 CGCTGTCGCC GGCCTGCTCC TCAAGCACGG CCCCAACAGT GAAGTAGCTG
 7801 ATTGTCATCA GCGCATTGAC GGCGTCCCCG GCCGAAAAAC CCGCCTCGCA
 7851 GAGGAAGCGA AGCTGCGCGT CGGCCGTTTC CATCTGCGGT GCGCCCGGTC
 7901 GCGTGCCGGC ATGGATGCGC GCGCCATCGC GGTAGGCGAG CAGCGCCTGC
 7951 CTGAAGCTGC GGGCATTCCC GATCAGAAAT GAGCGCCAGT CGTCGTCGGC
 8001 TCTCGGCACC GAATGCGTAT GATTCTCCGC CAGCATGGCT TCGGCCAGTG
 8051 CGTCGAGCAG CGCCCGCTTG TTCCTGAAGT GCCAGTAAAG CGCCGGCTGC
 8101 TGAACCCCCA ACCGTTCCGC CAGTTTGCGT GTCGTCAGAC CGTCTACGCC
 8151 GACCTCGTTC AACAGGTCCA GGGCGGCACG GATCACTGTA TTCGGCTGCA
 8201 ACTTTGTCAT GCTTGACACT TTATCACTGA TAAACATAAT ATGTCCACCA
 8251 ACTTATCAGT GATAAAGAAT CCGCGCGTTC AATCGGACCA GCGGAGGCTG
 8301 GTCCGGAGGC CAGACaTGAA ACCCAACATA CCCCTGATCG TAATTCTGAG
 8351 CACTGTCGCG CTCGACGCTG TCGGCATCGG CCTGATTATG CCGGTGCTGC
 8401 CGGGCCTCCT GCGCGATCTG GTTCACTCGA ACGACGTCAC CGCCCACTAT
 8451 GGCATTCTGC TGGCGCTGTA TGCGTTGGTG CAATTTGCCT GCGCACCTGT

 8501 GCTGGGCGCG CTGTCGGATC GTTTCGGGCG GCGGCCAATC TTGCTCGTCT
 8551 CGCTGGCCGG CGCCACTGTC GACTACGCCA TCATGGCGAC AGCGCCTTTC
 8601 CTTTGGGTTC TCTATATCGG GCGGATCGTG GCCGGCATCA CCGGGGCGAC
 8651 TGGGGCGGTA GCCGGCGCTT ATATTGCCGA TATCACTGAT GGCGATGAGC
 8701 GCGCGCGGCA CTTCGGCTTC ATGAGCGCCT GTTTCGGGTT CGGGATGGTC
 8751 GCGGGACCTG TGCTCGGTGG GCTGATGGGC GGTTTCTCCC CCCACGCTCC
 8801 GTTCTTCGCC GCGGCAGCCT TGAACGGCCT CAATTTCCTG ACGGGCTGTT
 8851 TCCTTTTGCC GGAGTCGCAC AAAGGCGAAC GCCGGCCGTT ACGCCGGGAG
 8901 GCTCTCAACC CGCTCGCTTC GTTCCGGTGG GCCCGGGGCA TGACCGTCGT
 8951 CGCCGCCCTG ATGGCGGTCT TCTTCATCAT GCAACTTGTC GGACAGGTGC
 9001 CGGCCGCGCT TTGGGTCATT TTCGGCGAGG ATCGCTTTCA CTGGGACGCG
 9051 ACCACGATCG GCATTTCGCT TGCCGCATTT GGCATTCTGC ATTCACTCGC
 9101 CCAGGCAATG ATCACCGGCC CTGTAGCCGC CCGGCTCGGC GAAAGGCGGG
 9151 CACTCATGCT CGGAATGATT GCCGACGGCA CAGGCTACAT CCTGCTTGCC
 9201 TTCGCGACAC GGGGATGGAT GGCGTTCCCG ATCATGGTCC TGCTTGCTTC
 9251 GGGTGGCATC GGAATGCCGG CGCTGCAAGC AATGTTGTCC AGGCAGGTGG
 9301 ATGAGGAACG TCAGGGGCAG CTGCAAGGCT CACTGGCGGC GCTCACCAGC
 9351 CTGACCTCGA TCGTCGGACC CCTCCTCTTC ACGGCGATCT ATGCGGCTTC
 9401 TATAACAACG TGGAACGGGT GGGCATGGAT TGCAGGCGCT GCCCTCTACT
 9451 TGCTCTGCCT GCCGGCGCTG CGTCGCGGGC TTTGGAGCGG CGCAGGGCAA
 9501 CGAGCCGATC GCTGATCGTG GAAACGATAG GCCTATGCCA TGCGGGTCAA
 9551 GGCGACTTCC GGCAAGCTAT ACGCGCCCTA GGAGTGCGGT TGGAACGTTG
 9601 GCCCAGCCAG ATACTCCCGA TCACGAGCAG GACGCCGATG ATTTGAAGCG
 9651 CACTCAGCGT CTGATCCAAG AACAACCATC CTAGCAACAC GGCGGTCCCC
 9701 GGGCTGAGAA AGCCCAGTAA GGAAACAACT GTAGGTTCGA GTCGCGAGAT
 9751 CCCCCGGAAC CAAAGGAAGT AGGTTAAACC CGCTCCGATC AGGCCGAGCC
 9801 ACGCCAGGCC GAGAACATTG GTTCCTGTAG GCATCGGGAT TGGCGGATCA
 9851 AACACTAAAG CTACTGGAAC GAGCAGAAGT CCTCCGGCCG CCAGTTGCCA
 9901 GGCGGTAAAG GTGAGCAGAG GCACGGGAGG TTGCCACTTG CGGGTCAGCA
 9951 CGGTTCCGAA CGCCATGGAA ACCGCCCCCG CCAGGCCCGC TGCGACGCCG
10001 ACAGGATCTA GCGCTGCGTT TGGTGTCAAC ACCAACAGCG CCACGCCCGC
10051 AGTTCCGCAA ATAGCCCCCA GGACCGCCAT CAATCGTATC GGGCTACCTA
10101 GCAGAGCGGC AGAGATGAAC ACGACCATCA GCGGCTGCAC AGCGCCTACC
10151 GTCGCCGCGA CCCCGCCCGG CAGGCGGTAG ACCGAAATAA ACAACAAGCT
```

-continued

```
10201 CCAGAATAGC GAAATATTAA GTGCGCCGAG GATGAAGATG CGCATCCACC
10251 AGATTCCCGT TGGAATCTGT CGGACGATCA TCACGAGCAA TAAACCCGCC
10301 GGCAACGCCC GCAGCAGCAT ACCGGCGACC CCTCGGCCTC GCTGTTCGGG
10351 CTCCACGAAA ACGCCGGACA GATGCGCCTT GTGAGCGTCC TTGGGGCCGT
10401 CCTCCTGTTT GAAGACCGAC AGCCCAATGA TCTCGCCGTC GATGTAGGCG
10451 CCGAATGCCA CGGCATCTCG CAACCGTTCA GCGAACGCCT CCATGGGCTT
10501 TTTCTCCTCG TGCTCGTAAA CGGACCCGAA CATCTCTGGA GCTTTCTTCA
10551 GGGCCGACAA TCGGATCTCG CGGAAATCCT GCACGTCGGC CGCTCCAAGC
10601 CGTCGAATCT GAGCCTTAAT CACAATTGTC AATTTTAATC CTCTGTTTAT
10651 CGGCAGTTCG TAGAGCGCGC CGTGCGTCCC GAGCGATACT GAGCGAAGCA
10701 AGTGCGTCGA GCAGTGCCCG CTTGTTCCTG AAATGCCAGT AAAGCGCTGG
10751 CTGCTGAACC CCCAGCCGGA ACTGACCCCA CAAGGCCCTA GCGTTTGCAA
10801 TGCACCAGGT CATCATTGAC CCAGGCGTGT TCCACCAGGC CGCTGCCTCG
10851 CAACTCTTCG CAGGCTTCGC CGACCTGCTC GCGCCACTTC TTCACGCGGG
10901 TGGAATCCGA TCCGCACATG AGGCGGAAGG TTTCCAGCTT GAGCGGGTAC
10951 GGCTCCCGGT GCGAGCTGAA ATAGTCGAAC ATCCGTCGGG CCGTCGGCGA
11001 CAGCTTGCGG TACTTCTCCC ATATGAATTT CGTGTAGTGG TCGCCAGCAA
11051 ACAGCACGAC GATTTCCTCG TCGATCAGGA CCTGGCAACG GGACGTTTTC
11101 TTGCCACGGT CCAGGACGCG GAAGCGGTGC AGCAGCGACA CCGATTCCAG
11151 GTGCCCAACG CGGTCGGACG TGAAGCCCAT CGCCGTCGCC TGTAGGCGCG
11201 ACAGGCATTC CTCGGCCTTC GTGTAATACC GGCCATTGAT CGACCAGCCC
11251 AGGTCCTGGC AAAGCTCGTA GAACGTGAAG GTGATCGGCT CGCCGATAGG
11301 GGTGCGCTTC GCGTACTCCA ACACCTGCTG CCACACCAGT TCGTCATCGT
11351 CGGCCCGCAG CTCGACGCCG GTGTAGGTGA TCTTCACGTC CTTGTTGACG
11401 TGGAAAATGA CCTTGTTTTG CAGCGCCTCG CGCGGGATTT TCTTGTTGCG
11451 CGTGGTGAAC AGGGCAGAGC GGGCCGTGTC GTTTGGCATC GCTCGCATCG
11501 TGTCCGGCCA CGGCGCAATA TCGAACAAGG AAAGCTGCAT TTCCTTGATC
11551 TGCTGCTTCG TGTGTTTCAG CAACGCGGCC TGCTTGGCCT CGCTGACCTG
11601 TTTTGCCAGG TCCTCGCCGG CGGTTTTTCG CTTCTTGGTC GTCATAGTTC
11651 CTCGCGTGTC GATGGTCATC GACTTCGCCA AACCTGCCGC CTCCTGTTCG
11701 AGACGACGCG AACGCTCCAC GGCGGCCGAT GGCGCGGGCA GGGCAGGGGG
11751 AGCCAGTTGC ACGCTGTCGC GCTCGATCTT GGCCGTAGCT TGCTGGACCA
11801 TCGAGCCGAC GGACTGGAAG GTTTCGCGGG GCGCACGCAT GACGGTGCGG
11851 CTTGCGATGG TTTCGGCATC CTCGGCGGAA AACCCCGCGT CGATCAGTTC

11901 TTGCCTGTAT GCCTTCCGGT CAAACGTCCG ATTCATTCAC CCTCCTTGCG
11951 GGATTGCCCC GACTCACGCC GGGGCAATGT GCCCTTATTC CTGATTTGAC
12001 CCGCCTGGTG CCTTGGTGTC CAGATAATCC ACCTTATCGG CAATGAAGTC
12051 GGTCCCGTAG ACCGTCTGGC CGTCCTTCTC GTACTTGGTA TTCCGAATCT
12101 TGCCCTGCAC GAATACCAGC GACCCCTTGC CCAAATACTT GCCGTGGGCC
12151 TCGGCCTGAG AGCCAAAACA CTTGATGCGG AAGAAGTCGG TGCGCTCCTG
12201 CTTGTCGCCG GTCGTGGCCG CGCCAACCTT TGCGATCCGC AAGCGCGCGG
12251 TCGCCATCTT CACGCTGGAA CAGTACGTCG AGGCGGGCAT CATGACCCGC
12301 GAGCAATACG AGGTCATTAA AAGCGCCGTG ATTGATGATA TAGCGGCCCG
12351 GCTGCTCCTG GTTCTCGCGC ACCGAAATGG GTGACTTCAC CCCGCGCTCT
12401 TTGATCGTGG CACCGATTTC CGCGATGCTC TCCGGGGAAA AGCCGGGGTT
12451 GTCGGCCGTC CGCGGCTGAT GCGGATCTTC GTCGATCAGG TCCAGGTCCA
12501 GCTCGATAGG GCCGGAACCG CCCTGAGACG CCGCAGGAGC GTCCAGGAGG
12551 CTCGACAGGT CGCCGATGCT ATCCAACCCC AGGCCGGACG GCTGCGCCGC
12601 GCCTGCGGCT TCCTGAGCGG CCGCAGCGGT GTTTTTCTTG GTGGTCTTGG
12651 CTTGAGCCGC AGTCATTGGG AAATCTCCAT CTTCGTGAAC ACGTAATCAG
12701 CCAGGGCGCG AACCTCTTTC GATGCCTTGC GCGCGGCCGT TTTCTTGATC
12751 TTCCAGACCG GCACACCGGA TGCGAGGGCA TCGGCGATGC TGCTGCGCAG
12801 GCCAACGGTG GCCGGAATCA TCATCTTGGG GTACGCGGCC AGCAGCTCGG
12851 CTTGGTGGCG CGCGTGGCGC GGATTCCGCG CATCGACCTT GCTGGGCACC
12901 ATGCCAAGGA ATTGCAGCTT GGCGTTCTTC TGGCGCACGT TCGCAATGGT
12951 CGTGACCATC TTCTTGATGC CCTGGATGCT GTACGCCTCA AGCTCGATGG
13001 GGGACAGCAC ATAGTCGGCC GCGAAGAGGG CGGCCGCCAG GCCGACGCCA
13051 AGGGTCGGGG CCGTGTCGAT CAGGCACACG TCGAAGCCTT GGTTCGCCAG
13101 GGCCTTGATG TTCGCCCCGA ACAGCTCGCG GGCGTCGTCC AGCGACAGCC
13151 GTTCGGCGTT CGCCAGTACC GGGTTGGACT CGATGAGGGC GAGGCGCGCG
13201 GCCTGGCCGT CGCCGGCTGC GGGTGCGGTT TCGGTCCAGC CGCCGGCAGG
13251 GACAGCGCCG AACAGCTTGC TTGCATGCAG GCCGGTAGCA AAGTCCTTGA
13301 GCGTGTAGGA CGCATTGCCC TGGGGGTCCA GGTCGATCAC GGCAACCCGC
13351 AAGCCGCGCT CGAAAAAGTC GAAGGCAAGA TGCACAAGGG TCGAAGTCTT
13401 GCCGACGCCG CCTTTCTGGT TGGCCGTGAC CAAAGTTTTC ATCGTTTGGT
13451 TTCCTGTTTT TTCTTGGCGT CCGCTTCCCA CTTCCGGACG ATGTACGCCT
13501 GATGTTCCGG CAGAACCGCC GTTACCCGCG CGTACCCCTC GGGCAAGTTC
13551 TTGTCCTCGA ACGCGGCCCA CACGCGATGC ACCGCTTGCG ACACTGCGCC

13601 CCTGGTCAGT CCCAGCGACG TTGCGAACGT CGCCTGTGGC TTCCCATCGA
13651 CTAAGACGCC CCGCGCTATC TCGATGGTCT GCTGCCCCAC TTCCAGCCCC
13701 TGGATCGCCT CCTGGAACTG GCTTTCGGTA AGCCGTTTCT TCATGGATAA
13751 CACCCATAAT TTGCTCCGCG CCTTGGTTGA ACATAGCGGT GACAGCCGCC
13801 AGCACATGAG AGAAGTTTAG CTAAACATTT CTCGCACGTC AACACCTTTA
13851 GCCGCTAAAA CTCGTCCTTG GCGTAACAAA ACAAAAGCCC GGAAACCGGG
13901 CTTTCGTCTC TTGCCGCTTA TGGCTCTGCA CCCGGCTCCA TCACCAACAG
13951 GTCGCGCACG CGCTTCACTC GGTTGCGGAT CGACACTGCC AGCCCAACAA
14001 AGCCGGTTGC CGCCGCCGCC AGGATCGCGC CGATGATGCC GGCCACACCG
14051 GCCATCGCCC ACCAGGTCGC CGCCTTCCGG TTCCATTCCT GCTGGTACTG
14101 CTTCGCAATG CTGGACCTCG GCTCACCATA GGCTGACCGC TCGATGGCGT
14151 ATGCCGCTTC TCCCCTTGGC GTAAAACCCA GCGCCGCAGG CGGCATTGCC
14201 ATGCTGCCCG CCGCTTTCCC GACCACGACG CGCGCACCAG GCTTGCGGTC
14251 CAGACCTTCG GCCACGGCGA GCTGCGCAAG GACATAATCA GCCGCCGACT
14301 TGGCTCCACG CGCCTCGATC AGCTCTTGCA CTCGCGCGAA ATCCTTGGCC
14351 TCCACGGCCG CCATGAATCG CGCACGCGGC GAAGGCTCCG CAGGGCCG
```

-continued

Broad-host-range expression vector with N-terminal 7 ± His tag followed by cleavage site for Tobacco Etch Virus protease

pRKHTTEVPL1Dpuf.seq Length: 14391

```
   1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGAGGATCG CCATGCACCA CCACCACCAC CACCACGAGA ACCTGTACTT
2601 CCAGTCCGCT AGCTAGGGAA GATCTTGAGG CCGGCCCTCC GTCGCGGGCG
2651 GCACCCACGC CCGCatCGAT TCCAAGGTTC AGCCATTGAG ACGGCTCCGC
2701 TTCGCGCGCA AGCGCGGGTT GGGCCGACTG CAAGCGGAGA GGGAAGCATG
2751 GCACTGCTCA GCTTCGAGCG AAAATATCGC GTGCCGGGGG GCACGCTGGT
2801 CGGCGGAAAC CTGTTCGACT TCTGGGTCGG CCCTTTCTAT GTCGGCTTCT
2851 TCGGGGTTGC GACGTTTTTC TTCGCGGCCC TGGGTATCAT TCTGATTGCC
2901 TGGAGTGCCG TACTCCAGGG TACCTGGAAC CCCCAACTCA TCTCTGTCTA
2951 CCCGCCGGCC CTTGAATATG GCCTGGGAGG TGCACCCCTC GCAAAAGGCG
3001 GGCTGTGGCA GATCATCACG ATCTGCGCCA CTGGTGCCTT CGTCAGCTGG
3051 GCGCTGCGCG AAGTCGAAAT CTGCCGTAAG CTGGGCATCG GGTACCACAT
3101 CCCGTTCGCC TTCGCGTTCG CCATCCTGGC CTACCTGACG CTGGTGCTGT
3151 TCCGCCCGGT GATGATGGGC GCCTGGGGCT ATGCCTTCCC CTACGGGATC
3201 TGGACGCACC TCGACTGGGT GTCGAACACG GGCTACACCT ACGGCAACTT
3251 CCACTACAAC CCTGCCCACA TGATCGCCAT CTCGTTCTTC TTCACGAACG
3301 CGCTGGCTCT GGCGCTGCAC GGCGCCCTTG TGCTCTCCGC GGCCAACCCC
3351 GAGAAGGGCA AGGAAATGCG GACGCCGGAT CACGAGGATA CGTTCTTCCG
```

-continued

```
3401 CGATCTGGTC GGCTACTCGA TCGGGACGCT CGGCATCCAC CGCCTCGGCC
3451 TGCTGCTCTC GCTGAGCGCC GTCTTCTTCA GCGCCCTCTG CATGATCATT
3501 ACCGGCACCA TCTGGTTCGA TCAGTGGGTC GACTGGTGGC AATGGTGGGT
3551 GAAGCTGCCG TGGTGGGCGA ACATCCCGGG AGGCATCAAT GGCTGAGTAT
3601 CAGAACATCT TCTCCCAGGT CCAGGTCCGC GGACCGGCCG ACCTGGGGAT
3651 GACCGAAGAC GTCAACCTGG CCAACCGTTC GGGCGTCGGT CCCTTCTCGA
3701 CCCTGCTCGG CTGGTTCGGC AACGCCCAGC TCGGCCCGAT CTATCTCGGC
3751 TCGCTCGGCG TCCTGTCCCT CTTCTCGGGC CTGATGTGGT TCTTCACCAT
3801 CGGGATCTGG TTCTGGTATC AGGCGGGCTG GAACCCGGCC GTCTTCCTGC
3851 GCGACCTGTT CTTCTTCTCG CTCGAGCCGC CGGCACCCGA ATACGGTCTG
3901 TCCTTCGCGG CTCCGCTGAA GGAAGGCGGG CTGTGGCTGA TCGCGTCGTT
3951 CTTCATGTTC GTCGCGGTCT GGTCCTGGTG GGGCCGCACC TATCTCCGCG
4001 CTCAGGCGCT GGGCATGGGC AAGCACACCG CCTGGGCGTT CCTCTCGGCC
4051 ATCTGGCTGT GGATGGTGCT GGGCTTCATC CGTCCGATCC TCATGGGGTC
4101 CTGGTCGGAA GCGGTTCCCT ACGGCATCTT CTCGCACCTC GACTGGACGA
4151 ACAACTTCTC GCTCGTCCAC GGCAACCTGT TCTACAACCC CTTCCACGGT
4201 CTCTCGATCG CCTTCCTCTA CGGGTCGGCC CTGCTCTTCG CGATGCACGG
4251 TGCGACCATC CTCGCGGTCT CCCGCTTCGG CGGCGAGCGC GAGCTGGAGC
4301 AGATCGCCGA CCGCGGGACG GCAGCGGAGC GGGCCGCCCT CTTCTGGCGC
4351 TGGACCATGG GTTTCAACGC CACGATGGAA GGCATCCACC GCTGGGCCAT
4401 CTGGATGGCG GTCCTCGTGA CCCTCACCGG CGGCATCGGg ATCCTGCTCT
4451 CGGGCACGGT CGTGGACAAC TGGTACGTCT GGGGCCAGAA CCACGGCATG
4501 GCGCCGCTGA ACTGAGGAGC GATCACAATG GCTGACAAGA CCATCTTCAA
4551 CGATCACCTC AACACCAATC CGAAGACCAA CCTTCGCCTC TGGGTCGCTT
4601 TCCAGATGAT GAAGGGTGCG GGCTGGGCTG GCGGCGTGTT CTTCGGGACG
4651 CTCCTTCTCA TCGGGTTCTT CCGGGTGGTC GGGCGGATGC TTCCGATCCA
4701 GGAGAACCAG GCTCCGGCGC CGAACATCAC CGGCGCTCTG GAGACCGGGA
4751 TCGAGCTGAT CAAGCATCTC GTCTGAGACA AGTCTCGGGG CAGGGCGGCG
4801 CGAGGCCGCC CGCTCCTCCA AGTCCGGGCC ATATCGCCGG CCCGGGTCCG
4851 GGGCGACACC ACAGCCCGGT TCCCTTCCTG TTGGCGACAG GGACCTGGTG
4901 CCGTGTGGAA GACCGCACGG CACCTTTTTG ACATTCACGG GAGGCTCTGA
4951 TGACCAATCC CACCCCGCGA CCCGAAACCC CGCTTTTGGA TCGCGTCTGC
5001 TGCCCGGCCG ACATGAAGGC GCTGAGTGAC GCCGAACTGG AGCGGCTGGC
5051 CGACGAAGTG CGTTCCGAGG TCAGTGATAG GGGTAGTTTC TTATTTTAGG

5101 CAGTTTATAT GAAATTAAGA CATGCAGATG TCACAGTGGA TATTGAACTG
5151 GTCTCGAAAG CTCAATATCC CCCAAAGCAC AAGCACAAAC TTCGACATCA
5201 TGCAGAAGCG TTTCCCGAAc cgcgtcttcg acgtgggcat cgccgagcag
5251 catgccgtga ccttcgcggc cggcctcgCc gGGGccggGa tgaagccctt
5301 ctgcgcgatc tattcctcgt tcctgcaacg gggttacgac cagatcgccc
5351 atgacgtggc gctgcagaac cttcccgtcc gcttcgtgat cgaccgggcg
5401 gggctcgtgg gggccgatgg cgcgacccat gcgggggccT TCGACGTTGG
5451 CTTCATCACT TCGCTGCCCA ACATGACCGT GATGGCCGCG GCCGACGAGG
5501 CCGAGCTCAT CCACATGATC gcCACCGCCG TGGCCTTCGA CGAGGGCCCC
5551 ATCGCCTTcC GCTTCCCGCG GGGCGAGGGG GTGGGCGTCG AGATGCCCGA
5601 GCGCGGGACG GTGCTGGAGC CCGgCCGGGG CCGCGTGGTG CGCGAAGGGA
5651 Cggatgtcgc gatcctctcc ttCGGCGCGc ATCTGCAcGA GgccTGGcAG
5701 GCggcGAAAC TTCTCGAGGc cGAGGGGGTG AGCGTGACCG TGGCCGACgc
5751 ccgCTTctCg CgCCCGCTCg AcACGGGGCT CATCGACCAG CTCGTgcGCC
5801 ATCACGCGGC GCTGGTAACG GTGGAGCAGG GGGCcATGGG CGGCTTCGGC
5851 GCCCATGTCA TGCACTATCT CGCCAATTCC GGCGGCTTCG ACGGGGGCCT
5901 CGCGCTCCGG GTCATGACGC TGCCCGACCG CTTCATCGAG CAGGCGAGCC
5951 CCGAGgACAT GTATGCCGAT GCGGGGCTGC GGGCCGAGGA TATCAAGCTT
6001 GGCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT TATCCGCTCA
6051 CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTAA AGCCTGGGGT
6101 GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT CACTGCCCGC
6151 TTTCCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA ATCGGCCAAC
6201 GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCGGTCTT GCCTTGCTCG
6251 TCGGTGATGT ACTTCACCAG CTCCGCGAAG TCGCTCTTCT TGATGGAGCG
6301 CATGGGGACG TGCTTGGCAA TCACGCGCAC CCCCCGGCCG TTTTAGCGGC
6351 TAAAAAAGTC ATGGCTCTGC CCTCGGGCGG ACCACGCCCA TCATGACCTT
6401 GCCAAGCTCG TCCTGCTTCT CTTCGATCTT CGCCAGCAGG GCGAGGATCG
6451 TGGCATCACC GAACCGCGCC GTGCGCGGGT CGTCGGTGAG CCAGAGTTTC
6501 AGCAGGCCGC CCAGGCGGCC CAGGTCGCCA TTGATGCGGG CCAGCTCGCG
6551 GACGTGCTCA TAGTCCACGA CGCCCGTGAT TTTGTAGCCC TGGCCGACGG
6601 CCAGCAGGTA GGCCGACAGG CTCATGCCGG CCGCCGCCGC CTTTTCCTCA
6651 ATCGCTCTTC GTTCGTCTGG AAGGCAGTAC ACCTTGATAG GTGGGCTGCC
6701 CTTCCTGGTT GGCTTGGTTT CATCAGCCAT CCGCTTGCCC TCATCTGTTA
6751 CGCCGGCGGT AGCCGGCCAG CCTCGCAGAG CAGGATTCCC GTTGAGCACC
```

-continued

```
 6801 GCCAGGTGCG AATAAGGGAC AGTGAAGAAG GAACACCCGC TCGCGGGTGG
 6851 GCCTACTTCA CCTATCCTGC CCGGCTGACG CCGTTGGATA CACCAAGGAA
 6901 AGTCTACACG AACCCTTTGG CAAAATCCTG TATATCGTGC GAAAAAGGAT
 6951 GGATATACCG AAAAAATCGC TATAATGACC CCGAAGCAGG GTTATGCAGC
 7001 GGAAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG GTATCCGGTA
 7051 AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGGAAA
 7101 CGCCTGGTAT CTTTATAGTC CTGTCGGGTT TCGCCACCTC TGACTTGAGC
 7151 GTCGATTTTT GTGATGCTCG TCAGGGGGGC GGAGCCTATG GAAAAACGCC
 7201 AGCAACGCGG CCTTTTTACG GTTCCTGGCC TTTTGCTGGC CTTTTGCTCA
 7251 CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC CGTATTACCG
 7301 CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC CGAGCGCAGC
 7351 GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CAGAAGGCCG CCAGAGAGGC
 7401 CGAGCGCGGC CGTGAGGCTT GGACGCTAGG GCAGGGCATG AAAAAGCCCG
 7451 TAGCGGGCTG CTACGGGCGT CTGACGCGGT GGAAAGGGGG AGGGGATGTT
 7501 GTCTACATGG CTCTGCTGTA GTGAGTGGGT TGCGCTCCGG CAGCGGTCCT
 7551 GATCAATCGT CACCCTTTCT CGGTCCTTCA ACGTTCCTGA CAACGAGCCT
 7601 CCTTTTCGCC AATCCATCGA CAATCACCGC GAGTCCCTGC TCGAACGCTG
 7651 CGTCCGGACC GGCTTCGTCG AAGGCGTCTA TCGCGGCCCG CAACAGCGGC
 7701 GAGAGCGGAG CCTGTTCAAC GGTGCCGCCG CGCTCGCCGG CATCGCTGTC
 7751 GCCGGCCTGC TCCTCAAGCA CGGCCCCAAC AGTGAAGTAG CTGATTGTCA
 7801 TCAGCGCATT GACGGCGTCC CCGGCCGAAA AACCCGCCTC GCAGAGGAAG
 7851 CGAAGCTGCG CGTCGGCCGT TTCCATCTGC GGTGCGCCCG GTCGCGTGCC
 7901 GGCATGGATG CGCGCGCCAT CGCGGTAGGC GAGCAGCGCC TGCCTGAAGC
 7951 TGCGGGCATT CCCGATCAGA AATGAGCGCC AGTCGTCGTC GGCTCTCGGC
 8001 ACCGAATGCG TATGATTCTC CGCCAGCATG GCTTCGGCCA GTGCGTCGAG
 8051 CAGCGCCCGC TTGTTCCTGA AGTGCCAGTA AAGCGCCGGC TGCTGAACCC
 8101 CCAACCGTTC CGCCAGTTTG CGTGTCGTCA GACCGTCTAC GCCGACCTCG
 8151 TTCAACAGGT CCAGGGCGGC ACGGATCACT GTATTCGGCT GCAACTTTGT
 8201 CATGCTTGAC ACTTTATCAC TGATAAACAT AATATGTCCA CCAACTTATC
 8251 AGTGATAAAG AATCCGCGCG TTCAATCGGA CCAGCGGAGG CTGGTCCGGA
 8301 GGCCAGACaT GAAACCCAAC ATACCCCTGA TCGTAATTCT GAGCACTGTC
 8351 GCGCTCGACG CTGTCGGCAT CGGCCTGATT ATGCCGGTGC TGCCGGGCCT
 8401 CCTGCGCGAT CTGGTTCACT CGAACGACGT CACCGCCCAC TATGGCATTC
 8451 TGCTGGCGCT GTATGCGTTG GTGCAATTTG CCTGCGCACC TGTGCTGGGC

 8501 GCGCTGTCGG ATCGTTTCGG GCGGCGGCCA ATCTTGCTCG TCTCGCTGGC
 8551 CGGCGCCACT GTCGACTACG CCATCATGGC GACAGCGCCT TTCCTTTGGG
 8601 TTCTCTATAT CGGGCGGATC GTGGCCGGCA TCACCGGGGC GACTGGGGCG
 8651 GTAGCCGGCG CTTATATTGC CGATATCACT GATGGCGATG AGCGCGCGCG
 8701 GCACTTCGGC TTCATGAGCG CCTGTTTCGG GTTCGGGATG GTCGCGGGAC
 8751 CTGTGCTCGG TGGGCTGATG GGCGGTTTCT CCCCCCACGC TCCGTTCTTC
 8801 GCCGCGGCAG CCTTGAACGG CCTCAATTTC CTGACGGGCT GTTTCCTTTT
 8851 GCCGGAGTCG CACAAAGGCG AACGCCGGCC GTTACGCCGG GAGGCTCTCA
 8901 ACCCGCTCGC TTCGTTCCGG TGGGGCCGGG GCATGACCGT CGTCGCCGCC
 8951 CTGATGGCGG TCTTCTTCAT CATGCAACTT GTCGGACAGG TGCCGGCCGC
 9001 GCTTTGGGTC ATTTTCGGCG AGGATCGCTT TCACTGGGAC GCGACCACGA
 9051 TCGGCATTTC GCTTGCCGCA TTTGGCATTC TGCATTCACT CGCCCAGGCA
 9101 ATGATCACCG GCCCTGTAGC CGCCCGGCTC GGCGAAAGGC GGGCACTCAT
 9151 GCTCGGAATG ATTGCCGACG GCACAGGCTA CATCCTGCTT GCCTTCGCGA
 9201 CACGGGGATG GATGGCGTTC CCGATCATGG TCCTGCTTGC TTCGGGTGGC
 9251 ATCGGAATGC CGGCGCTGCA AGCAATGTTG TCCAGGCAGG TGGATGAGGA
 9301 ACGTCAGGGG CAGCTGCAAG GCTCACTGGC GGCGCTCACC AGCCTGACCT
 9351 CGATCGTCGG ACCCCTCCTC TTCACGGCGA TCTATGCGGC TTCTATAACA
 9401 ACGTGGAACG GGTGGGCATG GATTGCAGGC GCTGCCCTCT ACTTGCTCTG
 9451 CCTGCCGGCG CTGCGTCGCG GGCTTTGGAG CGGCGCAGGG CAACGAGCCG
 9501 ATCGCTGATC GTGGAAACGA TAGGCCTATG CCATGCGGGT CAAGGCGACT
 9551 TCCGGCAAGC TATACGCGCC CTAGGAGTGC GGTTGGAACG TTGGCCCAGC
 9601 CAGATACTCC CGATCACGAG CAGGACGCCG ATGATTTGAA GCGCACTCAG
 9651 CGTCTGATCC AAGAACAACC ATCCTAGCAA CACGGCGGTC CCCGGGCTGA
 9701 GAAAGCCCAG TAAGGAAACA ACTGTAGGTT CGAGTCGCGA GATCCCCCGG
 9751 AACCAAAGGA AGTAGGTTAA ACCCGCTCCG ATCAGGCCGA GCCACGCCAG
 9801 GCGGAGAACA TTGGTTCCTG TAGGCATCGG GATTGGCGGA TCAAACACTA
 9851 AAGCTACTGG AACGAGCAGA AGTCCTCCGG CCGCCAGTTG CCAGGCGGTA
 9901 AAGGTGAGCA GAGGCACGGG AGGTTGCCAC TTGCGGGTCA GCACGGTTCC
 9951 GAACGCCATG GAAACCGCCC CCGCCAGGCC CGCTGCGACG CCGACAGGAT
10001 CTAGCGCTGC GTTTGGTGTC AACACCAACA GCGCCACGCC CGCAGTTCCG
10051 CAAATAGCCC CCAGGACCGC CATCAATCGT ATCGGGCTAC CTAGCAGAGC
10101 GGCAGAGATG AACACGACCA TCAGCGGCTG CACAGCGCCT ACCGTCGCCG
10151 CGACCCCGCC CGGCAGGCGG TAGACCGAAA TAAACAACAA GCTCCAGAAT
```

-continued

```
10201 AGCGAAATAT TAAGTGCGCC GAGGATGAAG ATGCGCATCC ACCAGATTCC
10251 CGTTGGAATC TGTCGGACGA TCATCACGAG CAATAAACCC GCCGGCAACG
10301 CCCGCAGCAG CATACCGGCG ACCCCTCGGC CTCGCTGTTC GGGCTCCACG
10351 AAAACGCCGG ACAGATGCGC CTTGTGAGCG TCCTTGGGGC CGTCCTCCTG
10401 TTTGAAGACC GACAGCCCAA TGATCTCGCC GTCGATGTAG GCGCCGAATG
10451 CCACGGCATC TCGCAACCGT TCAGCGAACG CCTCCATGGG CGGGGGCTCC
10501 TCGTGCTCGT AAACGGACCC GAACATCTCT GGAGCTTTCT TCAGGGCCGA
10551 CAATCGGATC TCGCGGAAAT CCTGCACGTC GGCCGCTCCA AGCCGTCGAA
10601 TCTGAGCCTT AATCACAATT GTCAATTTTA ATCCTCTGTT TATCGGCAGT
10651 TCGTAGAGCG CGCCGTGCGT CCCGAGCGAT ACTGAGCGAA GCAAGTGCGT
10701 CGAGCAGTGC CCGCTTGTTC CTGAAATGCC AGTAAAGCGC TGGCTGCTGA
10751 ACCCCCAGCC GGAACTGACC CCACAAGGCC CTAGCGTTTG CAATGCACCA
10801 GGTCATCATT GACCCAGGCG TGTTCCACCA GGCCGCTGCC TCGCAACTCT
10851 TCGCAGGCTT CGCCGACCTG CTCGCGCCAC TTCTTCACGC GGGTGGAATC
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 TGCAGCAGCG ACACCGATTC CAGGTGCCCA
11151 ACGCGGTCGG ACGTGAAGCC CATCGCCGTC GCCTGTAGGC GCGACAGGCA
11201 TTCCTCGGCC TTCGTGTAAT ACCGGCCATT GATCGACCAG CCCAGGTCCT
11251 GGCAAAGCTC GTAGAACGTG AAGGTGATCG GCTCGCCGAT AGGGGTGCGC
11301 TTCGCGTACT CCAACACCTG CTGCCACACC AGTTCGTCAT CGTCGGCCCG
11351 CAGCTCGACG CCGGTGTAGG TGATCTTCAC GTCCTTGTTG ACGTGGAAAA
11401 TGACCTTGTT TTGCAGCGCC TCGCGCGGGA TTTTCTTGTT GCGCGTGGTG
11451 AACAGGGCAG AGCGGGCCGT GTCGTTTGGC ATCGCTCGCA TCGTGTCCGG
11501 CCACGGCGCA ATATCGAACA AGGAAAGCTG CATTTCCTTG ATCTGCTGCT
11551 TCGTGTGTTT CAGCAACGCG GCCTGCTTGG CCTCGCTGAC CTGTTTTGCC
11601 AGGTCCTCGC CGGCGGTTTT TCGCTTCTTG GTCGTCATAG TTCCTCGCGT
11651 GTCGATGGTC ATCGACTTCG CCAAACCTGC CGCCTCCTGT TCGAGACGAC
11701 GCGAACGCTC CACGGCGGCC GATGGCGCGG GCAGGGCAGG GGGAGCCAGT
11751 TGCACGCTGT CGCGCTCGAT CTTGGCCGTA GCTTGCTGGA CCATCGAGCC
11801 GACGGACTGG AAGGTTTCGC GGGGCGCACG CATGACGGTG CGGCTTGCGA
11851 TGGTTTCGGC ATCCTCGGCG GAAAACCCCG CGTCGATCAG TTCTTGCCTG

11901 TATGCCTTCC GGTCAAACGT CCGATTCATT CACCCTCCTT GCGGGATTGC
11951 CCCGACTCAC GCCGGGGCAA TGTGCCCTTA TTCCTGATTT GACCCGCCTG
12001 GTGCCTTGGT GTCCAGATAA TCCACCTTAT CGGCAATGAA GTCGGTCCCG
12051 TAGACCGTCT GGCCGTCCTT CTCGTACTTG GTATTCCGAA TCTTGCCCTG
12101 CACGAATACC AGCGACCCCT TGCCCAAATA CTTGCCGTGG GCCTCGGCCT
12151 GAGAGCCAAA ACACTTGATG CGGAAGAAGT CGGTGCGCTC CTGCTTGTCG
12201 CCGGTCGTGG CCGCGCCAAC CTTTGCGATC CGCAAGCGCG CGGTCGCCAT
12251 CTTCACGCTG GAACAGTACG TCGAGGCGGG CATCATGACC CGCGAGCAAT
12301 ACGAGGTCAT TAAAAGCGCC GTGATTGATG ATATAGCGGC CCGGCTGCTC
12351 CTGGTTCTCG CGCACCGAAA TGGGTGACTT CACCCCGCGC TCTTTGATCG
12401 TGGCACCGAT TTCCGCGATG CTCTCCGGGG AAAAGCCGGG GTTGTCGGCC
12451 GTCCGCGGCT GATGCGGATC TTCGTCGATC AGGTCCAGGT CCAGCTCGAT
12501 AGGGCCGGAA CCGCCCTGAG ACGCCGCAGG AGCGTCCAGG AGGCTCGACA
12551 GGTCGCCGAT GCTATCCAAC CCCAGGCCGG ACGGCTGCGC CGCGCCTGCG
12601 GCTTCCTGAG CGGCCGCAGC GGTGTTTTTC TTGGTGGTCT TGGCTTGAGC
12651 CGCAGTCATT GGGAAATCTC CATCTTCGTG AACACGTAAT CAGCCAGGGC
12701 GCGAACCTCT TTCGATGCCT TGCGCGCGGC CGTTTTCTTG ATCTTCCAGA
12751 CCGGCACACC GGATGCGAGG GCATCGGCGA TGCTGCTGCG CAGGCCAACG
12801 GTGGCCGGAA TCATCATCTT GGGGTACGCG GCCAGCAGCT CGGCTTGGTG
12851 GCGCGCGTGG CGCGGATTCC GCGCATCGAC CTTGCTGGGC ACCATGCCAA
12901 GGAATTGCAG CTTGGCGTTC TTCTGGCGCA CGTTCGCAAT GGTCGTGACC
12951 ATCTTCTTGA TGCCCTGGAT GCTGTACGCC TCAAGCTCGA TGGGGGACAG
13001 CACATAGTCG GCCGCGAAGA GGGCGGCCGC CAGGCCGACG CCAAGGGTCG
13051 GGGCCGTGTC GATCAGGCAC ACGTCGAAGC CTTGGTTCGC CAGGGCCTTG
13101 ATGTTCGCCC CGAACAGCTC GCGGGCGTCG TCCAGCGACA GCCGTTCGGC
13151 GTTCGCCAGT ACCGGGTTGG ACTCGATGAG GGCGAGGCGC GCGGCCTGGC
13201 CGTCGCCGGC TGCGGGTGCG GTTTCGGTCC AGCCGCCGGC AGGGACAGCG
13251 CCGAACAGCT TGCTTGCATG CAGGCCGGTA GCAAAGTCCT TGAGCGTGTA
13301 GGACGCATTG CCCTGGGGGT CCAGGTCGAT CACGGCAACC CGCAAGCCGC
13351 GCTCGAAAAA GTCGAAGGCA AGATGCACAA GGGTCGAAGT CTTGCCGACG
13401 CCGCCTTTCT GGTTGGCCGT GACCAAAGTT TTCATCGTTT GGTTTCCTGT
13451 TTTTTCTTGG CGTCCGCTTC CCACTTCCGG ACGATGTACG CCTGATGTTC
13501 CGGCAGAACC GCCGTTACCC GCGCGTACCC CTCGGGCAAG TTCTTGTCCT
13551 CGAACGCGGC CCACACGCGA TGCACCGCTT GCGACACTGC GCCCCTGGTC

13601 AGTCCCAGCG ACGTTGCGAA CGTCGCCTGT GGCTTCCCAT CGACTAAGAC
13651 GCCCCGCGCT ATCTCGATGG TCTGCTGCCC CACTTCCAGC CCCTGGATCG
13701 CCTCCTGGAA CTGGCTTTCG GTAAGCCGTT TCTTCATGGA TAACACCCAT
13751 AATTTGCTCC GCGCCTTGGT TGAACATAGC GGTGACAGCC GCCAGCACAT
13801 GAGAGAAGTT TAGCTAAACA TTTCTCGCAC GTCAACACCT TTAGCCGCTA
13851 AAACTCGTCC TTGGCGTAAC AAAACAAAAG CCCGGAAACC GGGCTTTCGT
13901 CTCTTGCCGC TTATGGCTCT GCACCCGGCT CCATCACCAA CAGGTCGCGC
13951 ACGCGCTTCA CTCGGTTGCG GATCGACACT GCCAGCCCAA CAAAGCCGGT
14001 TGCCGCCGCC GCCAGGATCG CGCCGATGAT GCCGGCCACA CCGGCCATCG
14051 CCCACCAGGT CGCCGCCTTC CGGTTCCATT CCTGCTGGTA CTGCTTCGCA
14101 ATGCTGGACC TCGGCTCACC ATAGGCTGAC CGCTCGATGG CGTATGCCGC
14151 TTCTCCCCTT GGCGTAAAAC CCAGCGCCGC AGGCGGCATT GCCATGCTGC
14201 CCGCCGCTTT CCCGACCACG ACGCGCGCAC CAGGCTTGCG GTCCAGACCT
14251 TCGGCCACGG CGAGCTGCGC AAGGACATAA TCAGCCGCCG ACTTGGCTCC
14301 ACGCGCCTCG ATCAGCTCTT GCACTCGCGC GAAATCCTTG GCCTCCACGG
14351 CCGCCATGAA TCGCGCACGC GGCGAAGGCT 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 ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCTGGCCCG CAAGAGGGGC

1701 TCGAAGACCG ACAACGAAGC GGTGCGCGGA CTGATCCGCC AGCACCGTCC
1751 GCTCGTGCTC ATGGGGTCGA TCAACGAGAA GATTTACCTT GCGGAACTGA
1801 AAGCCGGTCA CGGCCCGGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA CGGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTTGC CCTCGATGCC GGTGCCGAGA
2151 GGGTCGTGAG AGAGACTGTC GAAGCCCTGA CGTGGGCCGG GCTTCGGCGA
2201 GAGGAAGGGA GAGAACCAAT GAGCGATCAT GCCGTCAACA CGCCGGTCCA
2251 TGCCGCCAGG GCCCACGGGC ACCGAGCACC TCGTGCCGAG TTCTACGTCT
2301 ACTTCGCCGT CATTCTGCTG GGCGCCTTCC CGGTGGCCTT CGTGAGCTGG
2351 ATCGTCTCGA CGATCCGCCA CCGCAGGCTT CCCAAGCGCG GCCCCTTCGC
2401 GTCCGCCTGG TTCGATGCCA AGGCGATCAC GCCGCTGATT TTCCGCGCCT
2451 GACCGCAGGT CAGGTTGCGA CACGCCATTC GTCGTCTCCC CAAGGGGCGG
2501 CGGATTAATC GGGAGGGCAT GGTGCCTTAC CGTAACCCAC GCCACCAGTA
2551 GGCAGGAGGA ACACGTGTCG TCCGGTGGTC ACCACCACCA CCACCACCAC
2601 CACCACCACT AATAGGCCGG CCCTCCGTCG CGGGCGGCAC CCACGCCCGC
2651 atCGATTCCA AGGTTCAGCC ATTGAGACGG CTCCGCTTCG CGCGCAAGCG
2701 CGGGTTGGGC CGACTGCAAG CGGAGAGGGA AGCATGGCAC TGCTCAGCTT
2751 CGAGCGAAAA TATCGCGTGC CGGGGGGCAC GCTGGTCGGC GGAAACCTGT
2801 TCGACTTCTG GGTCGGCCCT TTCTATGTCG GCTTCTTCGG GGTTGCGACG
2851 TTTTTCTTCG CGGCCCTGGG TATCATTCTG ATTGCCTGGA GTGCCGTACT
2901 CCAGGGTACC TGGAACCCCC AACTCATCTC TGTCTACCCG CCGGCCCTTG
2951 AATATGGCCT GGGAGGTGCA CCCCTCGCAA AAGGCGGGCT GTGGCAGATC
3001 ATCACGATCT GCGCCACTGG TGCCTTCGTC AGCTGGGCGC TGCGCGAAGT
3051 CGAAATCTGC CGTAAGCTGG GCATCGGGTA CCACATCCCG TTCGCCTTCG
3101 CGTTCGCCAT CCTGGCCTAC CTGACGCTGG TGCTGTTCCG CCCGGTGATG
3151 ATGGGCGCCT GGGGCTATGC CTTCCCCTAC GGGATCTGGA CGCACCTCGA
3201 CTGGGTGTCG AACACGGGCT ACACCTACGG CAACTTCCAC TACAACCCTG
3251 CCCACATGAT CGCCATCTCG TTCTTCTTCA CGAACGCGCT GGCTCTGGCG
3301 CTGCACGGCG CCCTTGTGCT CTCCGCGGCC AACCCCGAGA AGGGCAAGGA
3351 AATGCGGACG CCGGATCACG AGGATACGTT CTTCCGCGAT CTGGTCGGCT
```

-continued

```
3401 ACTCGATCGG GACGCTCGGC ATCCACCGCC TCGGCCTGCT GCTCTCGCTG
3451 AGCGCCGTCT TCTTCAGCGC CCTCTGCATG ATCATTACCG GCACCATCTG
3501 GTTCGATCAG TGGGTCGACT GGTGGCAATG GTGGGTGAAG CTGCCGTGGT
3551 GGGCGAACAT CCCGGGAGGC ATCAATGGCT GAGTATCAGA ACATCTTCTC
3601 CCAGGTCCAG GTCCGCGGAC CGGCCGACCT GGGGATGACC GAAGACGTCA
3651 ACCTGGCCAA CCGTTCGGGC GTCGGTCCCT TCTCGACCCT GCTCGGCTGG
3701 TTCGGCAACG CCCAGCTCGG CCCGATCTAT CTCGGCTCGC TCGGCGTCCT
3751 GTCCTCCTTC TCGGGCCTGA TGTGGTTCTT CACCATCGGG ATCTGGTTCT
3801 GGTATCAGGC GGGCTGGAAC CCGGCCGTCT TCCTGCGCGA CCTGTTCTTC
3851 TTCTCGCTCG AGCCGCCGGC ACCCGAATAC GGTCTGTCCT TCGCGGCTCC
3901 GCTGAAGGAA GGCGGGCTGT GGCTGATCGC GTCGTTCTTC ATGTTCGTCG
3951 CGGTCTGGTC CTGGTGGGGC CGCACCTATC TCCGCGCTCA GGCGCTGGGC
4001 ATGGGCAAGC ACACCGCCTG GGCGTTCCTC TCGGCCATCT GGCTGTGGAT
4051 GGTGCTGGGC TTCATCCGTC CGATCCTCAT GGGGTCCTGG TCGGAAGCGG
4101 TTCCCTACGG CATCTTCTCG CACCTCGACT GGACGAACAA CTTCTCGCTC
4151 GTCCACGGCA ACCTGTTCTA CAACCCCTTC CACGGTCTCT CGATCGCCTT
4201 CCTCTACGGG TCGGCCCTGC TCTTCGCGAT GCACGGTGCG ACCATCCTCG
4251 CGGTCTCCCG CTTCGGCGGC GAGCGCGAGC TGGAGCAGAT CGCCGACCGC
4301 GGGACGGCAG CGGAGCGGGC CGCCCTCTTC TGGCGCTGGA CCATGGGTTT
4351 CAACGCCACG ATGGAAGGCA TCCACCGCTG GGCCATCTGG ATGGCGGTCC
4401 TCGTGACCCT CACCGGCGGC ATCGGgATCC TGCTCTCGGG CACGGTCGTG
4451 GACAACTGGT ACGTCTGGGG CCAGAACCAC GGCATGGCGC CGCTGAACTG
4501 AGGAGCGATC ACAATGGCTG ACAAGACCAT CTTCAACGAT CACCTCAACA
4551 CCAATCCGAA GACCAACCTT CGCCTCTGGG TCGCTTTCCA GATGATGAAG
4601 GGTGCGGGCT GGGCTGGCGG CGTGTTCTTC GGGACGCTCC TTCTCATCGG
4651 GTTCTTCCGG GTGGTCGGGC GGATGCTTCC GATCCAGGAG AACCAGGCTC
4701 CGGCGCCGAA CATCACCGGC GCTCTGGAGA CCGGGATCGA GCTGATCAAG
4751 CATCTCGTCT GAGACAAGTC TCGGGGCAGG GCGGCGCGAG GCCGCCCGCT
4801 CCTCCAAGTC CGGGCCATAT CGCCGGCCCG GGTCCGGGGC GACACCACAG
4851 CCCGGTTCCC TTCCTGTTGG CGACAGGGAC CTGGTGCCGT GTGGAAGACC
4901 GCACGGCACC CTTTTGACAT TCACGGGAGG CTCTGATGAC CAATCCCACC
4951 CCGCGACCCG AAACCCCGCT TTTGGATCGC GTCTGCTGCC CGGCCGACAT
5001 GAAGGCGCTG AGTGACGCCG AACTGGAGCG GCTGGCCGAC GAAGTGCGTT
5051 CCGAGGTCAG TGATAGGGGT AGTTTCTTAT TTTAGGCAGT TTATATGAAA

5101 TTAAGACATG CAGATGTCAC AGTGGATATT GAACTGGTCT CGAAAGCTCA
5151 ATATCCCCCA AAGCACAAGC ACAAACTTCG ACATCATGCA GAAGCGTTTC
5201 CCGAAccgcg tcttcgacgt gggcatcgcc gagcagcatg ccgtgacctt
5251 cgcggccggc ctcgCcgGGG ccggGatgaa gcccttctgc gcgatctatt
5301 cctcgttcct gcaacggggt tacgaccaga tcgcccatga cgtggcgctg
5351 cagaaccttc ccgtccgctt cgtgatcgac cgggcggggc tcgtgggggc
5401 cgatggcgcg acccatgcgg gggccTTCGA CGTTGGCTTC ATCACTTCGC
5451 TGCCCAACAT GACCGTGATG GCCGCGGCCG ACGAGGCCGA GCTCATCCAC
5501 ATGATCgcCA CCGCCGTGGC CTTCGACGAG GGCCCCATCG CCTTcCGCTT
5551 CCCGCGGGGC GAGGGGGTGG GCGTCGAGAT GCCCGAGCGC GGGACGGTGC
5601 TGGAGCCCGg CCGGGGCCGC GTGGTGCGCG AAGGGACgga tgtcgcgatc
5651 ctctccttCG GCGCGcATCT GCAcGAGgcc TTGcAGGCgg cGAAACTTCT
5701 CGAGGccGAG GGGGTGAGCG TGACCGTGGC CGACgcccgC TTctCgCgCC
5751 CGCTCgAcAC GGGGCTCATC GACCAGCTCG TgcGCCATCA CGCGGCGCTG
5801 GTAACGGTGG AGCAGGGGGC cATGGGCGGC TTCGGCGCCC ATGTCATGCA
5851 CTATCTCGCC AATTCCGGCG GCTTCGACGG GGGCCTCGCG CTCCGGGTCA
5901 TGACGCTGCC CGACCGCTTC ATCGAGCAGG CGAGCCCCGA GgACATGTAT
5951 GCCGATGCGG GGCTGCGGGC CGAGGATATC AAGCTTGGCG TAATCATGGT
6001 CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC
6051 ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG
6101 CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA
6151 ACCTGTCGTG CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC
6201 GGTTTGCGTA TTGGGCGCTC GGTCTTGCCT TGCTCGTCGG TGATGTACTT
6251 CACCAGCTCC GCGAAGTCGC TCTTCTTGAT GGAGCGCATG GGGACGTGCT
6301 TGGCAATCAC GCGCACCCCC CGGCCGTTTT AGCGGCTAAA AAAGTCATGG
6351 CTCTGCCCTC GGGCGGACCA CGCCCATCAT GACCTTGCCA AGCTCGTCCT
6401 GCTTCTCTTC GATCTTCGCC AGCAGGGCGA GGATCGTGGC ATCACCGAAC
6451 CGCGCCGTGC GCGGGTCGTC GGTGAGCCAG AGTTTCAGCA GGCCGCCCAG
6501 GCGGCCCAGG TCGCCATTGA TGCGGGCCAG CTCGCGGACG TGCTCATAGT
6551 CCACGACGCC CGTGATTTTG TAGCCCTGGC CGACGGCCAG CAGGTAGGCC
6601 GACAGGCTCA TGCCGGCCGA CGCCGCCTTT TCCTCAATCG CTCTTCGTTC
6651 GTCTGGAAGG CAGTACACCT TGATAGGTGG GCTGCCCTTC CTGGTTGGCT
6701 TGGTTTCATC AGCCATCCGC TTGCCCTCAT CTGTTACGCC GGCGGTAGCC
6751 GGCCAGCCTC GCAGAGCAGG ATTCCCGTTG AGCACCGCCA GGTGCGAATA
```

-continued

```
 6801 AGGGACAGTG AAGAAGGAAC ACCCGCTCGC GGGTGGGCCT ACTTCACCTA
 6851 TCCTGCCCGG CTGACGCCGT TGGATACACC AAGGAAAGTC TACACGAACC
 6901 CTTTGGCAAA ATCCTGTATA TCGTGCGAAA AAGGATGGAT ATACCGAAAA
 6951 AATCGCTATA ATGACCCCGA AGCAGGGTTA TGCAGCGGAA AAGCGCCACG
 7001 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCGG
 7051 AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT
 7101 ATAGTCCTGT CGGGTGTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA
 7151 TGCTCGTCAG GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT
 7201 TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG
 7251 CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT
 7301 GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
 7351 GGAAGCGGAA GAGCGCCAGA AGGCCGCCAG AGAGGCCGAG CGCGGCCGTG
 7401 AGGCTTGGAC GCTAGGGCAG GGCATGAAAA AGCCCGTAGC GGGCTGCTAC
 7451 GGGCGTCTGA CGCGGTGGAA AGGGGGAGGG GATGTTGTCT ACATGGCTCT
 7501 GCTGTAGTGA GTGGGTTGCG CTCCGGCAGC GGTCCTGATC AATCGTCACC
 7551 CTTTCTCGGT CCTTCAACGT TCCTGACAAC GAGCCTCCTT TTCGCCAATC
 7601 CATCGACAAT CACCGCGAGT CCCTGCTCGA ACGCTGCGTC CGGACCGGCT
 7651 TCGTCGAAGG CGTCTATCGC GGCCCGCAAC AGCGGCGAGA GCGGAGCCTG
 7701 TTCAACGGTG CCGCCGCGCT CGCCGGCATC GCTGTCGCCG GCCTGCTCCT
 7751 CAAGCACGGC CCCAACAGTG AAGTAGCTGA TTGTCATCAG CGCATTGACG
 7801 GCGTCCCCGG CCGAAAAACC CGCCTCGCAG AGGAAGCGAA GCTGCGCGTC
 7851 GGCCGTTTCC ATCTGCGGTG CGCCCGGTCG CGTGCCGGCA TGGATGCGCG
 7901 CGCCATCGCG GTAGGCGAGC AGCGCCTGCC TGAAGCTGCG GGCATTCCCG
 7951 ATCAGAAATG AGCGCCAGTC GTCGTCGGCT CTCGGCACCG AATGCGTATG
 8001 ATTCTCCGCC AGCATGGCTT CGGCCAGTGC GTCGAGCAGC GCCCGCTTGT
 8051 TCCTGAAGTG CCAGTAAAGC GCCGGCTGCT GAACCCCCAA CCGTTCCGCC
 8101 AGTTTGCGTG TCGTCAGACC GTCTACGCCG ACCTCGTTCA ACAGGTCCAG
 8151 GGCGGCACGG ATCACTGTAT TCGGCTGCAA CTTTGTCATG CTTGACACTT
 8201 TATCACTGAT AAACATAATA TGTCCACCAA CTTATCAGTG ATAAAGAATC
 8251 CGCGCGTTCA ATCGGACCAG CGGAGGCTGG TCCGGAGGCC AGACaTGAAA
 8301 CCCAACATAC CCCTGATCGT AATTCTGAGC ACTGTCGCGC TCGACGCTGT
 8351 CGGCATCGGC CTGATTATGC CGGTGCTGCC GGGCCTCCTG CGCGATCTGG
 8401 TTCACTCGAA CGACGTCACC GCCCACTATG GCATTCTGCT GGCGCTGTAT
 8451 GCGTTGGTGC AATTTGCCTG CGCACCTGTG CTGGGCGCGC TGTCGGATCG

 8501 TTTCGGGCGG CGGCCAATCT TGCTCGTCTC GCTGGCCGGC GCCACTGTCG
 8551 ACTACGCCAT CATGGCGACA GCGCCTTTCC TTTGGGTTCT CTATATCGGG
 8601 CGGATCGTGG CCGGCATCAC CGGGGCGACT GGGGCGGTAG CCGGCGCTTA
 8651 TATTGCCGAT ATCACTGATG GCGATGAGCG CGCGCGGCAC TTCGGCTTCA
 8701 TGAGCGCCTG TTTCGGGTTC GGGATGGTCG CGGGACCTGT GCTCGGTGGG
 8751 CTGATGGGCG GTTTCTCCCC CCACGCTCCG TTCTTCGCCG CGGCAGCCTT
 8801 GAACGGCCTC AATTTCCTGA CGGGCTGTTT CCTTTTGCCG GAGTCGCACA
 8851 AAGGCGAACG CCGGCCGTTA CGCCGGGAGG CTCTCAACCC GCTCGCTTCG
 8901 TTCCGGTGGG CCCGGGGCAT GACCGTCGTC GCCGCCCTGA TGGCGGTCTT
 8951 CTTCATCATG CAACTTGTCG GACAGGTGCC GGCCGCGCTT TGGGTCATTT
 9001 TCGGCGAGGA TCGCTTTCAC TGGGACGCGA CCACGATCGG CATTTCGCTT
 9051 GCCGCATTTG GCATTCTGCA TTCACTCGCC CAGGCAATGA TCACCGGCCC
 9101 TGTAGCCGCC CGGCTCGGCG AAAGGCGGGC ACTCATGCTC GGAATGATTG
 9151 CCGACGGCAC AGGCTACATC CTGCTTGCCT TCGCGACACG GGGATGGATG
 9201 GCGTTCCCGA TCATGGTCCT GCTTGCTTCG GGTGGCATCG GAATGCCGGC
 9251 GCTGCAAGCA ATGTTGTCCA GGCAGGTGGA TGAGGAACGT CAGGGGCAGC
 9301 TGCAAGGCTC ACTGGCGGCG CTCACCAGCC TGACCTCGAT CGTCGGACCC
 9351 CTCCTCTTCA CGGCGATCTA TGCGGCTTCT ATAACAACGT GGAACGGGTG
 9401 GGCATGGATT GCAGGCGCTG CCCTCTACTT GCTCTGCCTG 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 ACAACCATCC TAGCAACACG GCGGTCCCCG GGCTGAGAAA GCCCAGTAAG
 9701 GAAACAACTG TAGGTTCGAG TCGCGAGATC CCCCGGAACC AAAGGAAGTA
 9751 GGTTAAACCC GCTCCGATCA GGCCGAGCCA CGCCAGGCCG AGAACATTGG
 9801 TTCCTGTAGG CATCGGGATT GGCGGATCAA ACACTAAAGC TACTGGAACG
 9851 AGCAGAAGTC CTCCGGCCGC CAGTTGCCAG GCGGTAAAGG TGAGCAGAGG
 9901 CACGGGAGGT TGCCACTTGC GGGTCAGCAC GGTTCCGAAC GCCATGGAAA
 9951 CCGCCCCCGC CAGGCCCGCT GCGACGCCGA CAGGATCTAG CGCTGCGTTT
10001 GGTGTCAACA CCAACAGCGC CACGCCCGCA GTTCCGCAAA TAGCCCCCAG
10051 GACCGCCATC AATCGTATCG GGCTACCTAG CAGAGCGGCA GAGATGAACA
10101 CGACCATCAG CGGCTGCACA GCGCCTACCG TCGCCGCGAC CCCGCCCGGC
10151 AGGCGGTAGA CCGAAATAAA CAACAAGCTC CAGAATAGCG AAATATTAAG
```

-continued

```
10201 TGAGCCGAGG ATGAAGATGC GCATCCACCA GATTCCCGTT GGAATCTGTC
10251 GGACGATCAT CACGAGCAAT AAACCCGCCG GCAACGCCCG CAGCAGCATA
10301 CCGGCGACCC CTCGGCCTCG CTGTTCGGGC TCCACGAAAA CGCCGGACAG
10351 ATGCGCCTTG TGAGCGTCCT TGGGGCCGTC CTCCTGTTTG AAGACCGACA
10401 GCCCAATGAT CTCGCCGTCG ATGTAGGCGC CGAATGCCAC GGCATCTCGC
10451 AACCGTTCAG CGAACGCCTC CATGGGCTTT TTCTCCTCGT GCTCGTAAAC
10501 GGACCCGAAC ATCTCTGGAG CTTTCTTCAG GGCCGACAAT CGGATCTCGC
10551 GGAAATCCTG CACGTCGGCC GCTCCAAGCC GTCGAATCTG AGCCTTAATC
10601 ACAATTGTCA ATTTTAATCC TCTGTTTATC GGCAGTTCGT AGAGCGCGCC
10651 GTGCGTCCCG AGCGATACTG AGCGAAGCAA GTGCGTCGAG CAGTGCCCGC
10701 TTGTTCCTGA AATGCCAGTA AAGCGCTGGC TGCTGAACCC CCAGCCGGAA
10751 CTGACCCCAC AAGGCCCTAG CGTTTGCAAT GCACCAGGTC ATCATTGACC
10801 CAGGCGTGTT CCACCAGGCC GCTGCCTCGC AACTCTTCGC AGGCTTCGCC
10851 GACCTGCTCG CGCCACTTCT TCACGCGGGT GGAATCCGAT CCGCACATGA
10901 GGCGGAAGGT TTCCAGCTTG AGCGGGTACG GCTCCCGGTG CGAGCTGAAA
10951 TAGTCGAACA TCCGTCGGGC CGTCGGCGAC AGCTTGCGGT ACTTCTCCCA
11001 TATGAATTTC GTGTAGTGGT CGCCAGCAAA CAGCACGACG ATTTCCTCGT
11051 CGATCAGGAC CTGGCAACGG GACGTTTTCT TGCCACGGTC CAGGACGCGG
11101 AAGCGGTGCA GCAGCGACAC CGATTCCAGG TGCCCAACGC GGTCGGACGT
11151 GAAGCCCATC GCCGTCGCCT GTAGGCGCGA CAGGCATTCC TCGGCCTTCG
11201 TGTAATACCG GCCATTGATC GACCAGCCCA GGTCCTGGCA AAGCTCGTAG
11251 AACGTGAAGG TGATCGGCTC GCCGATAGGG GTGCGCTTCG CGTACTCCAA
11301 CACCTGCTGC CACACCAGTT CGTCATCGTC GGCCCGCAGC TCGACGCCGG
11351 TGTAGGTGAT CTTCACGTCC TTGTTGACGT GGAAAATGAC CTTGTTTTGC
11401 AGCGCCTCGC GCGGGATTTT CTTGTTGCGC GTGGTGAACA GGGCAGAGCG
11451 GGCCGTGTCG TTTGGCATCG CTCGCATCGT GTCCGGCCAC GGCGCAATAT
11501 CGAACAAGGA AAGCTGCATT TCCTTGATCT GCTGCTTCGT GTGTTTCAGC
11551 AACGCGGCCT GCTTGGCCTC GCTGACCTGT TTTGCCAGGT CCTCGCCGGC
11601 GGTTTTTCGC TTCTTGGTCG TCATAGTTCC TCGCGTGTCG ATGGTCATCG
11651 ACTTCGCCAA ACCTGCCGCC TCCTGTTCGA GACGACGCGA ACGCTCCACG
11701 GCGGCCGATG GCGCGGGCAG GGCAGGGGGA GCCAGTTGCA CGCTGTCGCG
11751 CTCGATCTTG GCCGTAGCTT GCTGGACCAT CGAGCCGACG GACTGGAAGG
11801 TTTCGCGGGG CGCACGCATG ACGGTGCGGC TTGCGATGGT TTCGGCATCC
11851 TCGGCGGAAA ACCCCGCGTC GATCAGTTCT TGCCTGTATG CCTTCCGGTC

11901 AAACGTCCGA TTCATTCACC CTCCTTGCGG GATTGCCCCG ACTCACGCCG
11951 GGGCAATGTG CCCTTATTCC TGATTTGACC CGCCTGGTGC CTTGGTGTCC
12001 AGATAATCCA CCTTATCGGC AATGAAGTCG GTCCCGTAGA CCGTCTGGCC
12051 GTCCTTCTCG TACTTGGTAT TCCGAATCTT GCCCTGCACG AATACCAGCG
12101 ACCCCTTGCC CAAATACTTG CCGTGGGCCT CGGCCTGAGA GCCAAAACAC
12151 TTGATGCGGA AGAAGTCGGT GCGCTCCTGC TTGTCGCCGG TCGTGGCCGC
12201 GCCAACCTTT GCGATCCGCA AGCGCGCGGT CGCCATCTTC ACGCTGGAAC
12251 AGTACGTCGA GGCGGGCATC ATGACCCGCG AGCAATACGA GGTCATTAAA
12301 AGCGCCGTGA TTGATGATAT AGCGGCCCGG CTGCTCCTGG TTCTCGCGCA
12351 CCGAAATGGG TGACTTCACC CCGCGCTCTT TGATCGTGGC ACCGATTTCC
12401 GCGATGCTCT CCGGGGAAAA GCCGGGGTTG TCGGCCGTCC GCGGCTGATG
12451 CGGATCTTCG TCGATCAGGT CCAGGTCCAG CTCGATAGGG CCGGAACCGC
12501 CCTGAGACGC CGCAGGAGCG TCCAGGAGGC TCGACAGGTC GCCGATGCTA
12551 TCCAACCCCA GGCCGGACGG CTGCGCCGCG CCTGCGGCTT CCTGAGCGGC
12601 CGCAGCGGTG TTTTTCTTGG TGGTCTTGGC TTGAGCCGCA GTCATTGGGA
12651 AATCTCCATC TTCGTGAACA CGTAATCAGC CAGGGCGCGA ACCTCTTTCG
12701 ATGCCTTGCG CGCGGCCGTT TTCTTGATCT TCCAGACCGG CACACCGGAT
12751 GCGAGGGCAT CGGCGATGCT GCTGCGCAGG CCAACGGTGG CCGGAATCAT
12801 CATCTTGGGG TACGCGGCCA GCAGCTCGGC TTGGTGGCGC GCGTGGCGCG
12851 GATTCCGCGC ATCGACCTTG CTGGGCACCA TGCCAAGGAA TTGCAGCTTG
12901 GCGTTCTTCT GGCGCACGTT CGCAATGGTC GTGACCATCT TCTTGATGCC
12951 CTGGATGCTG TACGCCTCAA GCTCGATGGG GGACAGCACA TAGTCGGCCG
13001 CGAAGAGGGC GGCCGCCAGG CCGACGCCAA GGGTCGGGGC CGTGTCGATC
13051 AGGCACACGT CGAAGCCTTG GTTCGCCAGG GCCTTGATGT TCGCCCCGAA
13101 CAGCTCGCGG GCGTCGTCCA GCGACAGCCG TTCGGCGTTC GCCAGTACCG
13151 GGTTGGACTC GATGAGGGCG AGGCGCGCGG CCTGGCCGTC GCCGGCTGCG
13201 GGTGCGGTTT CGGTCCAGCC GCCGGCAGGG ACAGCGCCGA ACAGCTTGCT
13251 TGCATGCAGG CCGGTAGCAA AGTCCTTGAG CGTGTAGGAC GCATTGCCCT
13301 GGGGGTCCAG GTCGATCACG GCAACCCGCA AGCCGCGCTC GAAAAAGTCG
13351 AAGGCAAGAT GCACAAGGGT CGAAGTCTTG CCGACGCCGC CTTTCTGGTT
13401 GGCCGTGACC AAAGTTTTCA TCGTTTGGTT TCCTGTTTTT TCTTGGCGTC
13451 CGCTTCCCAC TTCCGGACGA TGTACGCCTG ATGTTCCGGC AGAACCGCCG
13501 TTACCCGCGC GTACCCCTCG GGCAAGTTCT TGTCCTCGAA CGCGGCCCAC
13551 ACGCGATGCA CCGCTTGCGA CACTGCGCCC CTGGTCAGTC CCAGCGACGT

13601 TGCGAACGTC GCCTGTGGCT TCCCATCGAC TAAGACGCCC CGCGCTATCT
13651 CGATGGTCTG CTGCCCCACT TCCAGCCCCT GGATCGCCTC CTGGAACTGG
13701 CTTTCGGTAA GCCGTTTCTT CATGGATAAC ACCCATAATT TGCTCCGCGC
13751 CTTGGTTGAA CATAGCGGTG ACAGCCGCCA GCACATGAGA GAAGTTTAGC
13801 TAAACATTTC TCGCACGTCA ACACCTTTAG CCGCTAAAAC TCGTCCTTGG
13851 CGTAACAAAA CAAAAGCCCG GAAACCGGGC TTTCGTCTCT TGCCGCTTAT
13901 GGCTCTGCAC CCGGCTCCAT CACCAACAGG TCGCGCACGC GCTTCACTCG
13951 GTTGCGGATC GACACTGCCA GCCCAACAAA GCCGGTTGCC GCCGCCGCCA
14001 GGATCGCGCC GATGATGCCG GCCACACCGG CCATCGCCCA CCAGGTCGCC
14051 GCCTTCCGGT TCCATTCCTG CTGGTACTGC TTCGCAATGC TGGACCTCGG
14101 CTCACCATAG GCTGACCGCT CGATGGCGTA TGCCGCTTCT CCCCTTGGCG
14151 TAAAACCCAG CGCCGCAGGC GGCATTGCCA TGCTGCCCGC CGCTTTCCCG
14201 ACCACGACGC GCGCACCAGG CTTGCGGTCC AGACCTTCGG CCACGGCGAG
14251 CTGCGCAAGG ACATAATCAG CCGCCGACTT GGCTCCACGC GCCTCGATCA
14301 GCTCTTGCAC TCGCGCGAAA TCCTTGGCCT CCACGGCCGC CATGAATCGC
14351 GCACGCGGCG AAGGCTCCGC AGGGCCG
```

-continued

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

featuring C-terminal 13 ± His tag

pRKLIC2HT13Dpuf.seq Length: 14386

```
   1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATCTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCTGGCCCG CAAGAGGGGC

1701 TCGAAGACCG ACAACGAAGC GGTGCGCGGA CTGATCCGCC AGCACCGTCC
1751 GCTCGTGCTC ATGGGGTCGA TCAACGAGAA GATTTACCTT GCGGAACTGA
1801 AAGCCGGTCA CGGCCCGGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA CGGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTTGC CCTCGATGCC GGTGCCGAGA
2151 GGGTCGTGAG AGAGACTGTC GAAGCCCTGA CGTGGGCCGG GCTTCGGCGA
2201 GAGGAAGGGA GAGAACCAAT GAGCGATCAT GCCGTCAACA CGCCGGTCCA
2251 TGCCGCCAGG GCCCACGGGC ACCGAGCACC TCGTGCCGAG TTCTACGTCT
2301 ACTTCGCCGT CATTCTGCTG GGCGCCTTCC CGGTGGCCTT CGTGAGCTGG
2351 ATCGTCTCGA CGATCCGCCA CCGCAGGCTT CCCAAGCGCG GCCCCTTCGC
2401 GTCCGCCTGG TTCGATGCCA AGGCGATCAC GCCGCTGATT TTCCGCGCCT
2451 GACCGCAGGT CAGGTTGCGA CACGCCATTC GTCGTCTCCC CAAGGGGCGG
2501 CGGATTAATC GGGAGGGCAT GGTGCCTTAC CGTAACCCAC GCCACCAGTA
2551 GGCAGGAGGA ACACGTGTCG TCCGGTGGTG ACCACGACCA GCACCACCAC
2601 CACGACCACC ACGACCACTA ATAGGCCGGC CCTCCGTCGC GGGCGGCACC
2651 CACGCCCGCa tCGATTCCAA GGTTCAGCCA TTGAGACGGC TCCGCTTCGC
2701 GCGCAAGCGC GGGTTGGGCC GACTGCAAGC GGAGAGGGAA GCATGGCACT
2751 GCTCAGCTTC GAGCGAAAAT ATCGCGTGCC GGGGGGCACG CTGGTCGGCG
2801 GAAACCTGTT CGACTTCTGG GTCGGCCCTT TCTATGTCGG CTTCTTCGGG
2851 GTTGCGACGT TTTTCTTCGC GGCCCTGGGT ATCATTCTGA TTGCCTGGAG
2901 TGCCGTACTC CAGGGTACCT GGAACCCCCA ACTCATCTCT GTCTACCCGC
2951 CGGCCCTTGA ATATGGCCTG GGAGGTGCAC CCCTCGCAAA AGGCGGGCTG
3001 TGGCAGATCA TCACGATCTG CGCCACTGGT GCCTTCGTCA GCTGGGCGCT
3051 GCGCGAAGTC GAAATCTGCC GTAAGCTGGG CATCGGGTAC CACATCCCGT
3101 TCGCCTTCGC GTTCGCCATC CTGGCCTACC TGACGCTGGT GCTGTTCCGC
3151 CCGGTGATGA TGGGCGCCTG GGGCTATGCC TTCCCCTACG GGATCTGGAC
3201 GCACCTCGAC TGGGTGTCGA ACACGGGCTA CACCTACGGC AACTTCCACT
3251 ACAACCCTGC CCACATGATC GCCATCTCGT TCTTCTTCAC GAACGCGCTG
3301 GCTCTGGCGC TGCACGGCGC CCTTGTGCTC TCCGCGGCCA ACCCCGAGAA
3351 GGGCAAGGAA ATGCGGACGC CGGATCACGA GGATACGTTC TTCCGCGATC
```

-continued

```
3401 TGGTCGGCTA CTCGATCGGG ACGCTCGGCA TCCACCGCCT CGGCCTGCTG
3451 CTCTCGCTGA GCGCCGTCTT CTTCAGCGCC CTCTGCATGA TCATTACCGG
3501 CACCATCTGG TTCGATCAGT GGGTCGACTG GTGGCAATGG TGGGTGAAGC
3551 TGCCGTGGTG GGCGAACATC CCGGGAGGCA TCAATGGCTG AGTATCAGAA
3601 CATCTTCTCC CAGGTCCAGG TCCGCGGACC GGCCGACCTG GGGATGACCG
3651 AAGACGTCAA CCTGGCCAAC CGTTCGGGCG TCGGTCCCTT CTCGACCCTG
3701 CTCGGCTGGT TCGGCAACGC CCAGCTCGGC CCGATCTATC TCGGCTCGCT
3751 CGGCGTCCTG TCCCTCTTCT CGGGCCTGAT GTGGTTCTTC ACCATCGGGA
3801 TCTGGTTCTG GTATCAGGCG GGCTGGAACC CGGCCGTCTT CCTGCGCGAC
3851 CTGTTCTTCT TCTCGCTCGA GCCGCCGGCA CCCGAATACG GTCTGTCCTT
3901 CGCGGCTCCG CTGAAGGAAG GCGGGCTGTG GCTGATCGCG TCGTTCTTCA
3951 TGTTCGTCGC GGTCTGGTCC TGGTGGGGCC GCACCTATCT CCGCGCTCAG
4001 GCGCTGGGCA TGGGCAAGCA CACCGCCTGG GCGTTCCTCT CGGCCATCTG
4051 GCTGTGGATG GTGCTGGGCT TCATCCGTCC GATCCTCATG GGGTCCTGGT
4101 CGGAAGCGGT TCCCTACGGC ATCTTCTCGC ACCTCGACTG GACGAACAAC
4151 TTCTCGCTCG TCCACGGCAA CCTGTTCTAC AACCCCTTCC ACGGTCTCTC
4201 GATCGCCTTC CTCTACGGGT CGGCCCTGCT CTTCGCGATG CACGGTGCGA
4251 CCATCCTCGC GGTCTCCCGC TTCGGCGGCG AGCGCGAGCT GGAGCAGATC
4301 GCCGACCGCG GGACGGCAGC GGAGCGGGCC GCCCTCTTCT GGCGCTGGAC
4351 CATGGGTTTC AACGCCACGA TGGAAGGCAT CCACCGCTGG GCCATCTGGA
4401 TGGCGGTCCT CGTGACCCTC ACCGGCGGCA TCGGgATCCT GCTCTCGGGC
4451 ACGGTCGTGG ACAACTGGTA CGTCTGGGGC CAGAACCACG GCATGGCGCC
4501 GCTGAACTGA GGAGCGATCA CAATGGCTGA CAAGACCATC TTCAACGATC
4551 ACCTCAACAC CAATCCGAAG ACCAACCTTC GCCTCTGGGT CGCTTTCCAG
4601 ATGATGAAGG GTGCGGGCTG GGCTGGCGGC GTGTTCTTCG GGACGCTCCT
4651 TCTCATCGGG TTCTTCCGGG TGGTCGGGCG GATGCTTCCG ATCCAGGAGA
4701 ACCAGGCTCC GGCGCCGAAC ATCACCGGCG CTCTGGAGAC CGGGATCGAG
4751 CTGATCAAGC ATCTCGTCTG AGACAAGTCT CGGGGCAGGG CGGCGCGAGG
4801 CCGCCCGCTC CTCCAAGTCC GGGCCATATC GCCGGCCCGG GTCCGGGGCG
4851 ACACCACAGC CCGGTTCCCT TCCTGTTGGC GACAGGGACC TGGTGCCGTG
4901 TGGAAGACCG CACGGCACCC TTTTGACATT CACGGGAGGC TCTGATGACC
4951 AATCCCACCC CGCGACCCGA AACCCCGCTT TTGGATCGCG TCTGCTGCCC
5001 GGCCGACATG AAGGCGCTGA GTGACGCCGA ACTGGAGCGG CTGGCCGACG
5051 AAGTGCGTTC CGAGGTCAGT GATAGGGGTA GTTTCTTATT TTAGGCAGTT

5101 TATATGAAAT TAAGACATGC AGATGTCACA GTGGATATTG AACTGGTCTC
5151 GAAAGCTCAA TATCCCCCAA AGCACAAGCA CAAACTTCGA CATCATGCAG
5201 AAGCGTTTCC CGAAccgcgt cttcgacgtg ggcatcgccg agcagcatgc
5251 cgtgaccttc gcggccggcc tcgCcgGGGc cggGatgaag cccttctgcg
5301 cgatctattc ctcgttcctg caacggggtt acgaccagat cgcccatgac
5351 gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct
5401 cgtgggggcc gatggcgcga cccatgcggg ggccTTCGAC GTTGGCTTCA
5451 TCACTTCGCT GCCCAACATG ACCGTGATGG CCGCGGCCGA CGAGGCCGAG
5501 CTCATCCACA TGATCgcCAC CGCCGTGGCC TTCGACGAGG GCCCCATCGC
5551 CTTcCGCTTC CCGCGGGGCG AGGGGGTGGG CGTCGAGATG CCCGAGCGCG
5601 GGACGGTGCT GGAGCCCGgC CGGGGCCGCG TGGTGCGCGA AGGGACggat
5651 gtcgcgatcc tctccttCGG CGCGcATCTG CAcGAGgccT TGcAGGCggc
5701 GAAACTTCTC GAGGccGAGG GGGTGAGCGT GACCGTGGCC GACgcccgCT
5751 TctCgCgCCC GCTCgAcACG GGGCTCATCG ACCAGCTCGT gcGCCATCAC
5801 GCGGCGCTGG TAACGGTGGA GCATGGGGCc ATGGGCGGCT TCGGCGCCCA
5851 TGTCATGCAC TATCTCGCCA ATTCCGGCGG CTTCGACGGG GGCCTCGCGC
5901 TCCGGGTCAT GACGCTGCCC GACCGCTTCA TCGAGCAGGC GAGCCCCGAG
5951 gACATGTATG CCGATGCGGG GCTGCGGGCC GAGGATATCA AGCTTGGCGT
6001 AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT
6051 CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT GGGGTGCCTA
6101 ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC
6151 AGTCGGGAAA CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG
6201 GGGAGAGGCG GTTTGCGTAT TGGGCGCTCG GTCTTGCCTT GCTCGTCGGT
6251 GATGTACTTC ACCAGCTCCG CGAAGTCGCT CTTCTTGATG GAGCGCATGG
6301 GGACGTGCTT GGCAATCACG CGCACCCCCC GGCCGTTTTA GCGGCTAAAA
6351 AAGTCATGGC TCTGCCCTCG GGCGGACCAC GCCCATCATG ACCTTGCCAA
6401 GCTCGTCCTG CTTCTCTTCG ATCTTCGCCA GCAGGGCGAG GATCGTGGCA
6451 TCACCGAACC GCGCCGTGCG CGGGTCGTCG GTGAGCCAGA GTTTCAGCAG
6501 GCCGCCCAGG CGGCCCAGGT CGCCATTGAT GCGGGCCAGC TCGCGGACGT
6551 GCTCATAGTC CACGACGCCC GTGATTTTGT AGCCCTGGCC GACGGCCAGC
6601 AGGTAGGCCG ACAGGCTCAT GCCGGCCGCC GCCGCCTTTT CCTCAATCGC
6651 TCTTCGTTCG TCTGGAAGGC AGTACACCTT GATAGGTGGG CTGCCCTTCC
6701 TGGTTGGCTT GGTTTCATCA GCCATCCGCT TGCCCTCATC TGTTACGCCG
6751 GCGGTAGCCG GCCAGCCTCG CAGAGCAGGA TTCCCGTTGA GCACCGCCAG
```

-continued

```
 6801 GTGCGAATAA GGGACAGTGA AGAAGGAACA CCCGCTCGCG GGTGGGCCTA
 6851 CTTCACCTAT CCTGCCCGGC TGACGCCGTT GGATACACCA AGGAAAGTCT
 6901 ACACGAACCC TTTGGCAAAA TCCTGTATAT CGTGCGAAAA AGGATGGATA
 6951 TACCGAAAAA ATCGCTATAA TGACCCCGAA GCAGGGTTAT GCAGCGGAAA
 7001 AGCGCCACGC TTCCCGAAGG GAGAAAGGCG GACAGGTATC CGGTAAGCGG
 7051 CAGGGTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAAACGCCT
 7101 GGTATCTTTA TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA
 7151 TTTTTGTGAT GCTCGTCAGG GGGGCGGAGC CTATGGAAAA ACGCCAGCAA
 7201 CGCGGCCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT
 7251 TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT TACCGCCTTT
 7301 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA ACGACCGAGC GCAGCGAGTC
 7351 AGTGAGCGAG GAAGCGGAAG AGCGCCAGAA GGCCGCCAGA GAGGCCGAGC
 7401 GCGGCCGTGA GGCTTGGACG CTAGGGCAGG GCATGAAAAA GCCCGTAGCG
 7451 GGCTGCTACG GGCGTCTGAC GCGGTGGAAA GGGGGAGGGG ATGTTGTCTA
 7501 CATGGCTCTG CTGTAGTGAG TGGGTTGCGC TCCGGCAGCG GTCCTGATCA
 7551 ATCGTCACCC TTTCTCGGTC CTTCAACGTT CCTGACAACG AGCCTCCTTT
 7601 TCGCCAATCC ATCGACAATC ACCGCGAGTC CCTGCTCGAA CGCTGCGTCC
 7651 GGACCGGCTT CGTCGAAGGC GTCTATCGCG GCCCGCAACA GCGGCGAGAG
 7701 CGGAGCCTGT TCAACGGTGC CGCCGCGCTC GCCGGCATCG CTGTCGCCGG
 7751 CCTGCTCCTC AAGCACGGCC CCAACAGTGA AGTAGCTGAT TGTCATCAGC
 7801 GCATTGACGG CGTCCCCGGC CGAAAAACCC GCCTCGCAGA GGAAGCGAAG
 7851 CTGCGCGTCG GCCGTTTCCA TCTGCGGTGC GCCCGGTCGC GTGCCGGCAT
 7901 GGATGCGCGC GCCATCGCGG TAGGCGAGCA GCGCCTGCCT GAAGCTGCGG
 7951 GCATTCCCGA TCAGAAATGA GCGCCAGTCG TCGTCGGCTC TCGGCACCGA
 8001 ATGCGTATGA TTCTCCGCCA GCATGGCTTC GGCCAGTGCG TCGAGCAGCG
 8051 CCCGCTTGTT CCTGAAGTGC CAGTAAAGCG CCGGCTGCTG AACCCCCAAC
 8101 CGTTCCGCCA GTTTGCGTGT CGTCAGACCG TCTACGCCGA CCTCGTTCAA
 8151 CAGGTCCAGG GCGGCACGGA TCACTGTATT CGGCTGCAAC TTTGTCATGC
 8201 TTGACACTTT ATCACTGATA AACATAATAT GTCCACCAAC TTATCAGTGA
 8251 TAAAGAATCC GCGCGTTCAA TCGGACCAGC GGAGGCTGGT CCGGAGGCCA
 8301 GACaTGAAAC CCAACATACC CCTGATCGTA ATTCTGAGCA CTGTCGCGCT
 8351 CGACGCTGTC GGCATCGGCC TGATTATGCC GGTGCTGCCG GGCCTCCTGC
 8401 GCGATCTGGT TCACTCGAAC GACGTCACCG CCCACTATGG CATTCTGCTG
 8451 GCGCTGTATG CGTTGGTGCA ATTTGCCTGC GCACCTGTGC TGGGCGCGCT

 8501 GTCGGATCGT TTCGGGCGGC GGCCAATCTT GCTCGTCTCG CTGGCCGGCG
 8551 CCACTGTCGA CTACGCCATC ATGGCGACAG CGCCTTTCCT TTGGGTTCTC
 8601 TATATCGGGC GGATCGTGGC CGGCATCACC GGGGCGACTG GGGCGGTAGC
 8651 CGGCGCTTAT ATTGCCGATA TCACTGATGG CGATGAGCGC GCGCGGCACT
 8701 TCGGCTTCAT GAGCGCCTGT TTCGGGTTCG GGATGGTCGC GGGACCTGTG
 8751 CTCGGTGGGC TGATGGGCGG TTTCTCCCCC CACGCTCCGT TCTTCGCCGC
 8801 GGCAGCCTTG AACGGCCTCA ATTTCCTGAC GGGCTGTTTC CTTTTGCCGG
 8851 AGTCGCACAA AGGCGAACGC CGGCCGTTAC GCCGGGAGGC TCTCAACCCG
 8901 CTCGCTTCGT TCCGGTGGGC CCGGGGCATG ACCGTCGTCG CCGCCCTGAT
 8951 GGCGGTCTTC TTCATCATGC AACTTGTCGG ACAGGTGCCG GCCGCGCTTT
 9001 GGGTCATTTT CGGCGAGGAT CGCTTTCACT GGGACGCGAC CACGATCGGC
 9051 ATTTCGCTTG CCGCATTTGG CATTCTGCAT TCACTCGCCC AGGCAATGAT
 9101 CACCGGCCCT GTAGCCGCCC GGCTCGGCGA AAGGCGGGCA CTCATGCTCG
 9151 GAATGATTGC CGACGGCACA GGCTACATCC TGCTTGCCTT CGCGACACGG
 9201 GGATGGATGG CGTTCCCGAT CATGGTCCTG CTTGCTTCGG GTGGCATCGG
 9251 AATGCCGGCG CTGCAAGCAA TGTTGTCCAG GCAGGTGGAT GAGGAACGTC
 9301 AGGGGCAGCT GCAAGGCTCA CTGGCGGCGC TCACCAGCCT GACCTCGATC
 9351 GTCGGACCCC TCCTCTTCAC GGCGATCTAT GCGGCTTCTA TAACAACGTG
 9401 GAACGGGTGG GCATGGATTG CAGGCGCTGC CCTCTACTTG CTCTGCCTGC
 9451 CGGCGCTGCG TCGCGGGCTT TGGAGCGGCG CAGGGCAACG AGCCGATCGC
 9501 TGATCGTGGA AACGATAGGC CTATGCCATG CGGGTCAAGG CGACTTCCGG
 9551 CAAGCTATAC GCGCCCTAGG AGTGCGGTTG GAACGTTGGC CCAGCCAGAT
 9601 ACTCCCGATC ACGAGCAGGA CGCCGATGAT TTGAAGCGCA CTCAGCGTCT
 9651 GATCCAAGAA CAACCATCCT AGCAACACGG CGGTCCCCGG GCTGAGAAAG
 9701 CCCAGTAAGG AAACAACTGT AGGTTCGAGT CGCGAGATCC CCCGGAACCA
 9751 AAGGAAGTAG GTTAAACCCG CTCCGATCAG GCCGAGCCAC GCCAGGCCGA
 9801 GAACATTGGT TCCTGTAGGC ATCGGGATTG GCGGATCAAA CACTAAAGCT
 9851 ACTGGAACGA GCAGAAGTCC TCCGGCCGCC AGTTGCCAGG CGGTAAAGGT
 9901 GAGCAGAGGC ACGGGAGGTT GCCACTTGCG GGTCAGCACG GTTCCGAACG
 9951 CCATGGAAAC CGCCCCCGCC AGGCCCGCTG CGACGCCGAC AGGATCTAGC
10001 GCTGCGTTTG GTGTCAACAC CAACAGCGCC ACGCCCGCAG TTCCGCAAAT
10051 AGCCCCCAGG ACCGCCATCA ATCGTATCGG GCTACCTAGC AGAGCGGCAG
10101 AGATGAACAC GACCATCAGC GGCTGCACAG CGCCTACCGT CGCCGCGACC
10151 CCGCCCGGCA GGCGGTAGAC CGAAATAAAC AACAAGCTCC AGAATAGCGA
```

-continued

```
10201 AATATTAAGT GCGCCGAGGA TGAAGATGCG CATCCACCAG ATTCCCGTTG
10251 GAATCTGTCG GACGATCATC ACGAGCAATA AACCCGCCGG CAACGCCCGC
10301 AGCAGCATAC CGGCGACCCC TCGGCCTCGC TGTTCGGGCT CCACGAAAAC
10351 GCCGGACAGA TGCGCCTTGT GAGCGTCCTT GGGGCCGTCC TCCTGTTTGA
10401 AGACCGACAG CCCAATGATC TCGCCGTCGA TGTAGGCGCC GAATGCCACG
10451 GCATCTCGCA ACCGTTCAGC GAACGCCTCC ATGGGCTTTT TCTCCTCGTG
10501 CTCGTAAACG GACCCGAACA TCTCTGGAGC TTTCTTCAGG GCCGACAATC
10551 GGATCTCGCG GAAATCCTGC ACGTCGGCCG CTCCAAGCCG TCGAATCTGA
10601 GCCTTAATCA CAATTGTCAA TTTTAATCCT CTGTTTATCG GCAGTTCGTA
10651 GAGCGCGCCG TGCGTCCCGA GCGATACTGA GCGAAGCAAG TGCGTCGAGC
10701 AGTGCCCGCT TGTTCCTGAA ATGCCAGTAA AGCGCTGGCT GCTGAACCCC
10751 CAGCCGGAAC TGACCCCACA AGGCCCTAGC GTTTGCAATG CACCAGGTCA
10801 TCATTGACCC AGGCGTGTTC CACCAGGCCG CTGCCTCGCA ACTCTTCGCA
10851 GGCTTCGCCG ACCTGCTCGC GCCACTTCTT CACGCGGGTG GAATCCGATC
10901 CGCACATGAG GCGGAAGGTT TCCAGCTTGA GCGGGTACGG CTCCCGGTGC
10951 GAGCTGAAAT AGTCGAACAT CCGTCGGGCC GTCGGCGACA GCTTGCGGTA
11001 CTTCTCCCAT ATGAATTTCG TGTAGTGGTC GCCAGCAAAC AGCACGACGA
11051 TTTCCTCGTC GATCAGGACC TGGCAACGGG ACGTTTTCTT GCCACGGTCC
11101 AGGACGCGGA AGCGGTGCAG CAGCGACACC GATTCCAGGT GCCCAACGCG
11151 GTCGGACGTG AAGCCCATCG CCGTCGCCTG TAGGCGCGAC AGGCATTCCT
11201 CGGCCTTCGT GTAATACCGG CCATTGATCG ACCAGCCCAG GTCCTGGCAA
11251 AGCTCGTAGA ACGTGAAGGT GATCGGCTCG CCGATAGGGG TGCGCTTCGC
11301 GTACTCCAAC ACCTGCTGCC ACACCAGTTC GTCATCGTCG GCCCGCAGCT
11351 CGACGCCGGT GTAGGTGATC TTCACGTCCT TGTTGACGTG GAAAATGACC
11401 TTGTTTTGCA GCGCCTCGCG CGGGATTTTC TTGTTGCGCG TGGTGAACAG
11451 GGCAGAGCGG GCCGTGTCGT TTGGCATCGC TCGCATCGTG TCCGGCCACG
11501 GCGCAATATC GAACAAGGAA AGCTGCATTT CCTTGATCTG CTGCTTCGTG
11551 TGTTTCAGCA ACGCGGCCTG CTTGGCCTCG CTGACCTGTT TTGCCAGGTC
11601 CTCGCCGGCG GTTTTTCGCT TCTTGGTCGT CATAGTTCCT CGCGTGTCGA
11651 TGGTCATCGA CTTCGCCAAA CCTGCCGCCT CCTGTTCGAG ACGACGCGAA
11701 CGCTCCACGG CGGCCGATGG CGCGGGCAGG GCAGGGGGAG CCAGTTGCAC
11751 GCTGTCGCGC TCGATCTTGG CCGTAGCTTG CTGGACCATC GAGCCGACGG
11801 ACTGGAAGGT TTCGCGGGGC GCACGCATGA CGGTGCGGCT TGCGATGGTT
11851 TCGGCATCCT CGGCGGAAAA CCCCGCGTCG ATCAGTTCTT GCCTGTATGC

11901 CTTCCGGTCA AACGTCCGAT TCATTCACCC TCCTTGCGGG ATTGCCCCGA
11951 CTCACGCCGG GGCAATGTGC CCTTATTCCT GATTTGACCC GCCTGGTGCC
12001 TTGGTGTCCA GATAATCCAC CTTATCGGCA ATGAAGTCGG TCCCGTAGAC
12051 CGTCTGGCCG TCCTTCTCGT ACTTGGTATT CCGAATCTTG CCCTGCACGA
12101 ATACCAGCGA CCCCTTGCCC AAATACTTGC CGTGGGCCTC GGCCTGAGAG
12151 CCAAAACACT TGATGCGGAA GAAGTCGGTG CGCTCCTGCT TGTCGCCGGT
12201 CGTGGCCGCG CCAACCTTTG CGATCCGCAA GCGCGCGGTC GCCATCTTCA
12251 CGCTGGAACA GTACGTCGAG GCGGGCATCA TGACCCGCGA GCAATACGAG
12301 GTCATTAAAA GCGCCGTGAT TGATGATATA GCGGCCCGGC TGCTCCTGGT
12351 TCTCGCGCAC CGAAATGGGT GACTTCACCC CGCGCTCTTT GATCGTGGCA
12401 CCGATTTCCG CGATGCTCTC CGGGGAAAAG CCGGGGTTGT CGGCCGTCCG
12451 CGGCTGATGC GGATCTTCGT CGATCAGGTC CAGGTCCAGC TCGATAGGGC
12501 CGGAACCGCC CTGAGACGCC GCAGGAGCGT CCAGGAGGCT CGACAGGTCG
12551 CCGATGCTAT CCAACCCCAG GCCGGACGGC TGCGCCGCGC CTGCGGCTTC
12601 CTGAGCGGCC GCAGCGGTGT TTTTCTTGGT GGTCTTGGCT TGAGCCGCAG
12651 TCATGGGGAA ATCTCCATCT TCGTGAACAC GTAATCAGCC AGGGCGCGAA
12701 CCTCTTTCGA TGCCTTGCGC GCGGCCGTTT TCTTGATCTT CCAGACCGGC
12751 ACACCGGATG CGAGGGCATC GGCGATGCTG CTGCGCAGGC CAACGGTGGC
12801 CGGAATCATC ATCTTGGGGT ACGCGGCCAG CAGCTCGGCT TGGTGGCGCG
12851 CGTGGCGCGG ATTCCGCGCA TCGACCTTGC TGGGCACCAT GCCAAGGAAT
12901 TGCAGCTTGG CGTTCTTCTG GCGCACGTTC GCAATGGTCG TGACCATCTT
12951 CTTGATGCCC TGGATGCTGT ACGCCTCAAG CTCGATGGGG GACAGCACAT
13001 AGTCGGCCGC GAAGAGGGCG GCCGCCAGGC CGACGCCAAG GGTCGGGGCC
13051 GTGTCGATCA GGCACACGTC GAAGCCTTGG TTCGCCAGGG CCTTGATGTT
13101 CGCCCCGAAC AGCTCGCGGG CGTCGTCCAG CGACAGCCGT TCGGCGTTCG
13151 CCAGTACCGG GTTGGACTCG ATGAGGGCGA GGCGCGCGGC CTGGCCGTCG
13201 CCGGCTGCGG GTGCGGTTTC GGTCCAGCCG CCGGCAGGGA CAGCGCCGAA
13251 CAGCTTGCTT GCATGCAGGC CGGTAGCAAA GTCCTTGAGC GTGTAGGACG
13301 CATTGCCCTG GGGGTCCAGG TCGATCACGG CAACCCGCAA GCCGCGCTCG
13351 AAAAAGTCGA AGGCAAGATG CACAAGGGTC GAAGTCTTGC CGACGCCGCC
13401 TTTCTGGTTG GCCGTGACCA AAGTTTTCAT CGTTTGGTTT CCTGTTTTTT
13451 CTTGGCGTCC GCTTCCCACT TCCGGACGAT GTACGCCTGA TGTTCCGGCA
13501 GAACCGCCGT TACCCGCGCG TACCCCTCGG GCAAGTTCTT GTCCTCGAAC
13551 GCGGCCCACA CGCGATGCAC CGCTTGCGAC ACTGCGCCCC TGGTCAGTCC

13601 CAGCGACGTT GCGAACGTCG CCTGTGGCTT CCCATCGACT AAGACGCCCC
13651 GCGCTATCTC GATGGTCTGC TGCCCCACTT CCAGCCCCTG GATCGCCTCC
13701 TGGAACTGGC TTTCGGTAAG CCGTTTCTTC ATGGATAACA CCCATAATTT
13751 GCTCCGCGCC TTGGTTGAAC ATAGCGGTGA CAGCCGCCAG CACATGAGAG
13801 AAGTTTAGCT AAACATTTCT CGCACGTCAA CACCTTTAGC TTCGTCTCTT
13851 CGTCCTTGGC GTAACAAAAC AAAAGCCCGG AAADDGGGCT TTCGTCTCTT
13901 GCCGCTTATG GCTCTGCACC CGGCTCCATC ACCAACAGGT CGCGCACGCG
13951 CTTCACTCGG TTGCGGATCG ACACTGCCAG CCCAACAAAG CCGGTTGCCG
14001 CCGCCGCCAG GATCGCGCCG ATGATGCCGG CCACACCGGC CATCGCCCAC
14051 CAGGTCGCCG CCTTCCGGTT CCATTCCTGC TGGTACTGCT TCGCAATGCT
14101 GGACCTCGGC TCACCATAGG CTGACCGCTC GATGGCGTAT GCCGCTTCTC
14151 CCCTTGGCGT AAAACCCAGC GCCGCAGGCG GCATTGCCAT GCTGCCCGCC
14201 GCTTTCCCGA CCACGACGCG CGCACCAGGC TTGCGGTCCA GACCTTCGGC
14251 CACGGCGAGC TGCGCAAGGA CATAATCAGC CGCCGACTTG GCTCCACGCG
14301 CCTCGATCAG CTCTTGCACT CGCGCGAAAT CCTTGGCCTC CACGGCCGCC
14351 ATGAATCGCG CACGCGGCGA 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 ACCTGGTCGC TGAATGTCGA
  51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
 101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
 151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
 201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
 251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
 301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
 351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
 401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
 451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
 501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
 551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
 601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
 651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
 701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
 751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
 801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
 851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
 901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
 951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCTGGCCCG CAAGAGGGGC

1701 TCGAAGACCG ACAACGAAGC GGTGCGCGGA CTGATCCGCC AGCACCGTCC
1751 GCTCGTGCTC ATGGGGTCGA TCAACGAGAA GATTTACCTT GCGGAACTGA
1801 AAGCCGGTCA CGGCCCGGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA CGGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTTGC CCTCGATGCC GGTGCCGAGA
2151 GGGTCGTGAG AGAGACTGTC GAAGCCCTGA CGTGGGCCGG GCTTCGGCGA
2201 GAGGAAGGGA GAGAACCAAT GAGCGATCAT GCCGTCAACA CGCCGGTCCA
2251 TGCCGCCAGG GCCCACGGGC ACCGAGCACC TCGTGCCGAG TTCTACGTCT
2301 ACTTCGCCGT CATTCTGCTG GGCGCCTTCC CGGTGGCCTT CGTGAGCTGG
2351 ATCGTCTCGA CGATCCGCCA CCGCAGGCTT CCCAAGCGCG GCCCCTTCGC
2401 GTCCGCCTGG TTCGATGCCA AGGCGATCAC GCCGCTGATT TTCCGCGCCT
2451 GACCGCAGGT CAGGTTGCGA CACGCCATTC GTCGTCTCCC CAAGGGGCGG
2501 CGGATTAATC GGGAGGGCAT GGTGCCTTAC CGTAACCCAC GCCACCAGTA
2551 GGCAGGAGGA ACACGTGTCG TCCGGTGGTC ACXXXXXXXX XXXXXXXXXX
2601 TAATAGGCCG GCCCTCCGTC GCGGGCGGCA CCCACGCCCG CatCGATTCC
2651 AAGGTTCAGC CATTGAGACG GCTCCGCTTC GCGCGCAAGC GCGGGTTGGG
2701 CCGACTGCAA GCGGAGAGGG AAGCATGGCA CTGCTCAGCT TCGAGCGAAA
2751 ATATCGCGTG CCGGGGGGCA CGCTGGTCGG CGGAAACCTG TTCGACTTCT
2801 GGGTCGGCCC TTTCTATGTC GGCTTCTTCG GGGTTGCGAC GTTTTTCTTC
2851 GCGGCCCTGG GTATCATTCT GATTGCCTGG AGTGCCGTAC TCCAGGGTAC
2901 CTGGAACCCC CAACTCATCT CTGTCTACCC GCCGGCCCTT GAATATGGCC
2951 TGGGAGGTGC ACCCCTCGCA AAAGGCGGGC TGTGGCAGAT CATCACGATC
3001 TGCGCCACTG GTGCCTTCGT CAGCTGGGCG CTGCGCGAAG TCGAAATCTG
3051 CCGTAAGCTG GGCATCGGGT ACCACATCCC GTTCGCCTTC GCGTTCGCCA
3101 TCCTGGCCTA CCTGACGCTG GTGCTGTTCC GCCCGGTGAT GATGGGCGCC
3151 TGGGGCTATG CCTTCCCCTA CGGGATCTGG ACGCACCTCG ACTGGGTGTC
3201 GAACACGGGC TACACCTACG GCAACTTCCA CTACAACCCT GCCCACATGA
3251 TCGCCATCTC GTTCTTCTTC ACGAACGCGC TGGCTCTGGC GCTGCACGGC
3301 GCCCTTGTGC TCTCCGCGGC CAACCCCGAG AAGGGCAAGG AAATGCGGAC
3351 GCCGGATCAC GAGGATACGT TCTTCCGCGA TCTGGTCGGC TACTCGATCG
```

-continued

```
3401 GGACGCTCGG CATCCACCGC CTCGGCCTGC TGCTCTCGCT GAGCGCCGTC
3451 TTCTTCAGCG CCCTCTGCAT GATCATTACC GGCACCATCT GGTTCGATCA
3501 GTGGGTCGAC TGGTGGCAAT GGTGGGTGAA GCTGCCGTGG TGGGCGAACA
3551 TCCCGGGAGG CATCAATGGC TGAGTATCAG AACATCTTCT CCCAGGTCCA
3601 GGTCCGCGGA CCGGCCGACC TGGGGATGAC CGAAGACGTC AACCTGGCCA
3651 ACCGTTCGGG CGTCGGTCCC TTCTCGACCC TGCTCGGCTG GTTCGGCAAC
3701 GCCCAGCTCG GCCCGATCTA TCTCGGCTCG CTCGGCGTCC TGTCCCTCTT
3751 CTCGGGCCTG ATGTGGTTCT TCACCATCGG GATCTGGTTC TGGTATCAGG
3801 CGGGCTGGAA CCCGGCCGTC TTCCTGCGCG ACCTGTTCTT CTTCTCGCTC
3851 GAGCCGCCGG CACCCGAATA CGGTCTGTCC TTCGCGGCTC CGCTGAAGGA
3901 AGGCGGGCTG TGGCTGATCG CGTCGTTCTT CATGTTCGTC GCGGTCTGGT
3951 CCTGGTGGGG CCGCACCTAT CTCCGCGCTC AGGCGCTGGG CATGGGCAAG
4001 CACACCGCCT GGGCGTTCCT CTCGGCCATC TGGCTGTGGA TGGTGCTGGG
4051 CTTCATCCGT CCGATCCTCA TGGGGTCCTG GTCGGAAGCG GTTCCCTACG
4101 GCATCTTCTC GCACCTCGAC TGGACGAACA ACTTCTCGCT CGTCCACGGC
4151 AACCTGTTCT ACAACCCCTT CCACGGTCTC TCGATCGCCT TCCTCTACGG
4201 GTCGGCCCTG CTCTTCGCGA TGCACGGTGC GACCATCCTC GCGGTCTCCC
4251 GCTTCGGCGG CGAGCGCGAG CTGGAGCAGA TCGCCGACCG CGGGACGGCA
4301 GCGGAGCGGG CCGCCCTCTT CTGGCGCTGG ACCATGGGTT TCAACGCCAC
4351 GATGGAAGGC ATCCACCGCT GGGCCATCTG GATGGCGGTC CTCGTGACCC
4401 TCACCGGCGG CATCGGgATC CTGCTCTCGG GCACGGTCGT GGACAACTGG
4451 TACGTCTGGG GCCAGAACCA CGGCATGGCG CCGCTGAACT GAGGAGCGAT
4501 CACAATGGCT GACAAGACCA TCTTCAACGA TCACCTCAAC ACCAATCCGA
4551 AGACCAACCT TCGCCTCTGG GTCGCTTTCC AGATGATGAA GGGTGCGGGC
4601 TGGGCTGGCG GCGTGTTCTT CGGGACGCTC CTTCTCATCG GGTTCTTCCG
4651 GGTGGTCGGG CGGATGCTTC CGATCCAGGA GAACCAGGCT CCGGCGCCGA
4701 ACATCACCGG CGCTCTGGAG ACCGGGATCG AGCTGATCAA GCATCTCGTC
4751 TGAGACAAGT CTCGGGGCAG GGCGGCGCGA GGCCGCCCGC TCCTCCAAGT
4801 CCGGGCCATA TCGCCGGCCC GGGTCCGGGG CGACACCACA GCCCGGTTCC
4851 CTTCCTGTTG GCGACAGGGA CCTGGTGCCG TGTGGAAGAC CGCACGGCAC
4901 CCTTTTGACA TTCACGGGAG GCTCTGATGA CCAATCCCAC CCCGCGACCC
4951 GAAACCCCGC TTTTGGATCG CGTCTGCTGC CCGGCCGACA TGAAGGCGCT
5001 GAGTGACGCC GAACTGGAGC GGCTGGCCGA CGAAGTGCGT TCCGAGGTCA
5051 GTGATAGGGG TAGTTTCTTA TTTTAGGCAG TTTATATGAA ATTAAGACAT

5101 GCAGATGTCA CAGTGGATAT TGAACTGGTC TCGAAAGCTC AATATCCCCC
5151 AAAGCACAAG CACAAACTTC GACATCATGC AGAAGCGTTT CCCGAAccgc
5201 gtcttcgacg tgggcatcgc cgagcagcat gccgtgacct tcgcggccgg
5251 cctcgCcgGG GccggGatga agcccttctg cgcgatctat tcctcgttcc
5301 tgcaacgggg ttacgaccag atcgcccatg acgtggcgct gcagaacctt
5351 cccgtccgct tcgtgatcga ccgggcgggg ctcgtggggg ccgatggcgc
5401 gacccatgcg ggggccTTCG ACGTTGGCTT CATCACTTCG CTGCCCAACA
5451 TGACCGTGAT GGCCGCGGCC GACGAGGCCG AGCTCATCCA CATGATCgcC
5501 ACCGCCGTGG CCTTCGACGA GGGCCCCATC GCCTTcCGCT TCCCGCGGGG
5551 CGAGGGGGTG GGCGTCGAGA TGCCCGAGCG CGGGACGGTG CTGGAGCCCG
5601 gCCGGGGCCG CGTGGTGCGC GAAGGGACgg atgtcgcgat cctctccttC
5651 GGCGCGcATC TGCAcGAGgc cTTGcAGGCg gcGAAACTTC TCGAGGccGA
5701 GGGGGTGAGC GTGACCGTGG CCGACgcccg CTTctCgCgC CCGCTCgAcA
5751 CGGGGCTCAT CGACCAGCTC GTgcGCCATC ACGCGGCGCT GGTAACGGTG
5801 GAGCAGGGGG CcATGGGCGG CTTCGGCGCC CATGTCATGC ACTATCTCGC
5851 CAATTCCGGC GGCTTCGACG GGGGCCTCGC GCTCCGGGTC ATGACGCTGC
5901 CCGACCGCTT CATCGAGCAG GCGAGCCCCG AGgACATGTA TGCCGATGCG
5951 GGGCTGCGGG CCGAGGATAT CAAGCTTGGC GTAATCATGG TCATAGCTGT
6001 TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC
6051 GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC
6101 ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT
6151 GCCAGCTGCA TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT
6201 ATTGGGCGCT CGGTCTTGCC TTGCTCGTCG GTGATGTACT TCACCAGCTC
6251 CGCGAAGTCG CTCTTCTTGA TGGAGCGCAT GGGGACGTGC TTGGCAATCA
6301 CGCGCACCCC CCGGCCGTTT TAGCGGCTAA AAAAGTCATG GCTCTGCCCT
6351 CGGGCGGACC ACGCCCATCA TGACCTTGCC AAGCTCGTCC TGCTTCTCTT
6401 CGATCTTCGC CAGCAGGGCG AGGATCGTGG CATCACCGAA CCGCGCCGTG
6451 CGCGGGTCGT CGGTGAGCCA GAGTTTCAGC AGGCCGCCCA GGCGGCCCAG
6501 GTCGCCATTG ATGCGGGCCA GCTCGCGGAC GTGCTCATAG TCCACGACGC
6551 CCGTGATTTT GTAGCCCTGG CCGACGGCCA GCAGGTAGGC CGACAGGCTC
6601 ATGCCGGCCG CCGCCGCCTT TTCCTCAATC GCTCTTCGTT CGTCTGGAAG
6651 GCAGTACACC TTGATAGGTG GGCTGCCCTT CCTGGTTGGC TTGGTTTCAT
6701 CAGCCATCCG CTTGCCCTCA TCTGTTACGC CGGCGGTAGC CGGCCAGCCT
6751 CGCAGAGCAG GATTCCCGTT GAGCACCGCC AGGTGCGAAT AAGGGACAGT
```

-continued

```
 6801 GAAGAAGGAA CACCCGCTCG CGGGTGGGCC TACTTCACCT ATCCTGCCCG
 6851 GCTGACGCCG TTGGATACAC CAAGGAAAGT CTACACGAAC CCTTTGGCAA
 6901 AATCCTGTAT ATCGTGCGAA AAAGGATGGA TATACCGAAA AAATCGCTAT
 6951 AATGACCCCG AAGCAGGGTT ATGCAGCGGA AAAGCGCCAC GCTTCCCGAA
 7001 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA
 7051 GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG
 7101 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA
 7151 GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT
 7201 CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC
 7251 CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT
 7301 CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA
 7351 AGAGCGCCAG AAGGCCGCCA GAGAGGCCGA GCGCGGCCGT GAGGCTTGGA
 7401 CGCTAGGGCA GGGCATGAAA AAGCCCGTAG CGGGCTGCTA CGGGCGTCTG
 7451 ACGCGGTGGA AAGGGGGAGG GGATGTTGTC TACATGGCTC TGCTGTAGTG
 7501 AGTGGGTTGC GCTCCGGCAG CGGTCCTGAT CAATCGTCAC CCTTTCTCGG
 7551 TCCTTCAACG TTCCTGACAA CGAGCCTCCT TTTCGCCAAT CCATCGACAA
 7601 TCACCGCGAG TCCCTGCTCG AACGCTGCGT CCGGACCGGC TTCGTCGAAG
 7651 GCGTCTATCG CGGCCCGCAA CAGCGGCGAG AGCGGAGCCT GTTCAACGGT
 7701 GCCGCCGCGC TCGCCGGCAT CGCTGTCGCC GGCCTGCTCC TCAAGCACGG
 7751 CCCCAACAGT GAAGTAGCTG ATTGTCATCA GCGCATTGAC GGCGTCCCCG
 7801 GCCGAAAAAC CCGCCTCGCA GAGGAAGCGA AGCTGCGCGT CGGCCGTTTC
 7851 CATCTGCGGT GCGCCCGGTC GCGTGCCGGC ATGGATGCGC GCGCCATCGC
 7901 GGTAGGCGAG CAGCGCCTGC CTGAAGCTGC GGGCATTCCC GATCAGAAAT
 7951 GAGCGCCAGT CGTCGTCGGC TCTCGGCACC GAATGCGTAT GATTCTCCGC
 8001 CAGCATGGCT TCGGCCAGTG CGTCGAGCAG CGCCCGCTTG TTCCTGAAGT
 8051 GCCAGTAAAG CGCCGGCTGC TGAACCCCCA ACCGTTCCGC CAGTTTGCGT
 8101 GTCGTCAGAC CGTCTACGCC GACCTCGTTC AACAGGTCCA GGGCGGCACG
 8151 GATCACTGTA TTCGGCTGCA ACTTTGTCAT GCTTGACACT TTATCACTGA
 8201 TAAACATAAT ATGTCCACCA ACTTATCAGT GATAAAGAAT CCGCGCGTTC
 8251 AATCGGACCA GCGGAGGCTG GTCCGGAGGC CAGACaTGAA ACCCAACATA
 8301 CCCCTGATCG TAATTCTGAG CACTGTCGCG CTCGACGCTG TCGGCATCGG
 8351 CCTGATTATG CCGGTGCTGC CGGGCCTCCT GCGCGATCTG GTTCACTCGA
 8401 ACGACGTCAC CGCCCACTAT GGCATTCTGC TGGCGCTGTA TGCGTTGGTG
 8451 CAATTTGCCT GCGCACCTGT GCTGGGCGCG CTGTCGGATC CTTTCGGGCG

 8501 GCGGCCAATC TTGCTCGTCT CGCTGGCCGG CGCCACTGTC GACTACGCCA
 8551 TCATGGCGAC AGCGCCTTTC CTTTGGGTTC TCTATATCGG GCGGATCGTG
 8601 GCCGGCATCA CCGGGGCGAC TGGGGCGGTA GCCGGCGCTT ATATTGCCGA
 8651 TATCACTGAT GGCGATGAGC GCGCGCGGCA CTTCGGCTTC ATGAGCGCCT
 8701 GTTTCGGGTT CGGGATGGTC GCGGGACCTG TGCTCGGTGG GCTGATGGGC
 8751 GGTTTCTCCC CCCACGCTCC GTTCTTCGCC GCGGCAGCCT TGAACGGCCT
 8801 CAATTTCCTG ACGGGCTGTT TCCTTTTGCC GGAGTCGCAC AAAGGCGAAC
 8851 GCCGGCCGTT ACGCCGGGAG GCTCTCAACC CGCTCGCTTC GTTCCGGTGG
 8901 GCCCGGGGCA TGACCGTCGT CGCCGCCCTG ATGGCGGTCT TCTTCATCAT
 8951 GCAACTTGTC GGACAGGTGC CGGCCGCGCT TTGGGTCATT TTCGGCGAGG
 9001 ATCGCTTTCA CTGGGACGCG ACCACGATCG GCATTTCGCT TGCCGCATTT
 9051 GGCATTCTGC ATTCACTCGC CCAGGCAATG ATCACCGGCC CTGTAGCCGC
 9101 CCGGCTCGGC GAAAGGCGGG CACTCATGCT CGGAATGATT GCCGACGGCA
 9151 CAGGCTACAT CCTGCTTGCC TTCGCGACAC GGGGATGGAT GGCGTTCCCG
 9201 ATCATGGTCC TGCTTGCTTC GGGTGGCATC GGAATGCCGG CGCTGCAAGC
 9251 AATGTTGTCC AGGCAGGTGG ATGAGGAACG TCAGGGGCAG CTGCAAGGCT
 9301 CACTGGCGGC GCTCACCAGC CTGACCTCGA TCGTCGGACC CCTCCTCTTC
 9351 ACGGCGATCT ATGCGGCTTC TATAACAACG TGGAACGGGT GGGCATGGAT
 9401 TGCAGGCGCT GCCCTCTACT TGCTCTGCCT GCCGGCGCTG CGTCGCGGGC
 9451 TTTGGAGCGG CGCAGGGCAA CGAGCCGATC GCTGATCGTG GAAACGATAG
 9501 GCCTATGCCA TGCGGGTCAA GGCGACTTCC GGCAAGCTAT ACGCGCCCTA
 9551 GGAGTGCGGT TGGAACGTTG GCCCAGCCAG ATACTCCCGA TCACGAGCAG
 9601 GACGCCGATG ATTTGAAGCG CACTCAGCGT CTGATCCAAG AACAACCATC
 9651 CTAGCAACAC GGCGGTCCCC GGGCTGAGAA AGCCCAGTAA GGAAACAACT
 9701 GTAGGTTCGA GTCGCGAGAT CCCCCGGAAC CAAAGGAAGT AGGTTAAACC
 9751 CGCTCCGATC AGGCCGAGCC ACGCCAGGCC GAGAACATTG GTTCCTGTAG
 9801 GCATCGGGAT TGGCGGATCA AACACTAAAG CTACTGGAAC GAGCAGAAGT
 9851 CCTCCGGCCG CCAGTTGCCA GGCGGTAAAG GTGAGCAGAG GCACGGGAGG
 9901 TTGCCACTTG CGGGTCAGCA CGGTTCCGAA CGCCATGGAA ACCGCCCCCG
 9951 CCAGGCCCGC TGCGACGCCG ACAGGATCTA GCGCTGCGTT TGGTGTCAAC
10001 ACCAACAGCG CCACGCCCGC AGTTCCGCAA ATAGCCCCCA GGACCGCCAT
10051 CAATCGTATC GGGCTACCTA GCAGAGCGGC AGAGATGAAC ACGACCATCA
10101 GCGGCTGCAC AGCGCCTACC GTCGCCGCGA CCCCGCCCGG CAGGCGGTAG
10151 ACCGAAATAA ACAACAAGCT CCAGAATAGC GAAATATTAA GTGCGCCGAG
```

-continued

```
10201 GATGAAGATG CGCATCCACC AGATTCCCGT TGGAATCTGT CGGACGATCA
10251 TCACGAGCAA TAAACCCGCC GGCAACGCCC GCAGCAGCAT ACCGGCGACC
10301 CCTCGGCCTC GCTGTTCGGG CTCCACGAAA ACGCCGGACA GATGCGCCTT
10351 GTGAGCGTCC TTGGGGCCGT CCTCCTGTTT GAAGACCGAC AGCCCAATGA
10401 TCTCGCCGTC GATGTAGGCG CCGAATGCCA CGGCATCTCG CAACCGTTCA
10451 GCGAACGCCT CCATGGGCTT TTTCTCCTCG TGCTCGTAAA CGGACCCGAA
10501 CATCTCTGGA GCTTTCTTCA GGGCCGACAA TCGGATCTCG CGGAAATCCT
10551 GCACGTCGGC CGCTCCAAGC CGTCGAATCT GAGCCTTAAT CACAATTGTC
10601 AATTTTAATC CTCTGTTTAT CGGCAGTTCG TAGAGCGCGC CGTGCGTCCC
10651 GAGCGATACT GAGCGAAGCA AGTGCGTCGA GCAGTGCCCG CTTGTTCCTG
10701 AAATGCCAGT AAAGCGCTGG CTGCTGAACC CCCAGCCGGA ACTGACCCCA
10751 CAAGGCCCTA GCGTTTGCAA TGCACCAGGT CATCATTGAC CCAGGCGTGT
10801 TCCACCAGGC CGCTGCCTCG CAACTCTTCG CAGGCTTCGC CGACCTGCTC
10851 GCGCCACTTC TTCACGCGGG TGGAATCCGA TCCGCACATG AGGCGGAAGG
10901 TTTCCAGCTT GAGCGGGTAC GGCTCCCGGT GCGAGCTGAA ATAGTCGAAC
10951 ATCCGTCGGG CCGTCGGCGA CAGCTTGCGG TACTTCTCCC ATATGAATTT
11001 CGTGTAGTGG TCGCCAGCAA ACAGCACGAC GATTTCCTCG TCGATCAGGA
11051 CCTGGCAACG GGACGTTTTC TTGCCACGGT CCAGGACGCG GAAGCGGTGC
11101 AGCAGCGACA CCGATTCCAG GTGCCCAACG CGGTCGGACG TGAAGCCCAT
11151 CGCCGTCGCC TGTAGGCGCG ACAGGCATTC CTCGGCCTTC GTGTAATACC
11201 GGCCATTGAT CGACCAGCCC AGGTCCTGGC AAAGCTCGTA GAACGTGAAG
11251 GTGATCGGCT CGCCGATAGG GGTGCGCTTC GCGTACTCCA ACACCTGCTG
11301 CCACACCAGT TCGTCATCGT CGGCCCGCAG CTCGACGCCG GTGTAGGTGA
11351 TCTTCACGTC CTTGTTGACG TGGAAAATGA CCTTGTTTTG CAGCGCCTCG
11401 CGCGGGATTT TCTTGTTGCG CGTGGTGAAC AGGGCAGAGC GGGCCGTGTC
11451 GTTTGGCATC GCTCGCATCG TGTCCGGCCA CGGCGCAATA TCGAACAAGG
11501 AAAGCTGCAT TTCCTTGATC TGCTGCTTCG TGTGTTTCAG CAACGCGGCC
11551 TGCTTGGCCT CGCTGACCTG TTTTGCCAGG TCCTCGCCGG CGGTTTTTCG
11601 CTTCTTGGTC GTCATAGTTC CTCGCGTGTC GATGGTCATC GACTTCGCCA
11651 AACCTGCCGC CTCCTGTTCG AGACGACGCG AACGCTCCAC GGCGGCCGAT
11701 GGCGCGGGCA GGGCAGGGGG AGCCAGTTGC ACGCTGTCGC GCTCGATCTT
11751 GGCCGTAGCT TGCTGGACCA TCGAGCCGAC GGACTGGAAG GTTTCGCGGG
11801 GCGCACGCAT GACGGTGCGG CTTGCGATGG TTTCGGCATC CTCGGCGGAA
11851 AACCCCGCGT CGATCAGTTC TTGCCTGTAT GCCTTCCGGT CAAACGTCCG

11901 ATTCATTCAC CCTCCTTGCG GGATTGCCCC GACTCACGCC GGGGCAATGT
11951 GCCCTTATTC CTGATTTGAC CCGCCTGGTG CCTTGGTGTC CAGATAATCC
12001 ACCTTATCGG CAATGAAGTC GGTCCCGTAG ACCGTCTGGC CGTCCTTCTC
12051 GTACTTGGTA TTCCGAATCT TGCCCTGCAC GAATACCAGC GACCCCTTGC
12101 CCAAATACTT GCCGTGGGCC TCGGCCTGAG AGCCAAAACA CTTGATGCGG
12151 AAGAAGTCGG TGCGCTCCTG CTTGTCGCCG GTCGTGGCCG CGCCAACCTT
12201 TGCGATCCGC AAGCGCGCGG TCGCCATCTT CACGCTGGAA CAGTACGTCG
12251 AGGCGGGCAT CATGACCCGC GAGCAATACG AGGTCATTAA AAGCGCCGTG
12301 ATTGATGATA TAGCGGCCCG GCTGCTCCTG GTTCTCGCGC ACCGAAATGG
12351 GTGACTTCAC CCCGCGCTCT TTGATCGTGG CACCGATTTC CGCGATGCTC
12401 TCCGGGGAAA AGCCGGGGTT GTCGGCCGTC CGCGGCTGAT GCGGATCTTC
12451 GTCGATCAGG TCCAGGTCCA GCTCGATAGG GCCGGAACCG CCCTGAGACG
12501 CCGCAGGAGC GTCCAGGAGG CTCGACAGGT CGCCGATGCT ATCCAACCCC
12551 AGGCCGGACG GCTGCGCCGC GCCTGCGGCT TCCTGAGCGG CCGCAGCGGT
12601 GTTTTTCTTG GTGGTCTTGG CTTGAGCCGC AGTCATTGGG AAATCTCCAT
12651 CTTCGTGAAC ACGTAATCAG CCAGGGCGCG AACCTCTTTC GATGCCTTGC
12701 GCGCGGCCGT TTTCTTGATC TTCCAGACCG GCACACCGGA TGCGAGGGCA
12751 TCGGCGATGC TGCTGCGCAG GCCAACGGTG GCCGGAATCA TCATCTTGGG
12801 GTACGCGGCC AGCAGCTCGG CTTGGTGGCG CGCGTGGCGC GGATTCCGCG
12851 CATCGACCTT GCTGGGCACC ATGCCAAGGA ATTGCAGCTT GGCGTTCTTC
12901 TGGCGCACGT TCGCAATGGT CGTGACCATC TTCTTGATGC CCTGGATGCT
12951 GTACGCCTCA AGCTCGATGG GGGACAGCAC ATAGTCGGCC GCGAAGAGGG
13001 CGGCCGCCAG GCCGACGCCA AGGGTCGGGG CCGTGTCGAT CAGGCACACG
13051 TCGAAGCCTT GGTTCGCCAG GGCCTTGATG TTCGCCCCGA ACAGCTCGCG
13101 GGCGTCGTCC AGCGACAGCC GTTCGGCGTT CGCCAGTACC GGGTTGGACT
13151 CGATGAGGGC GAGGCGCGCG GCCTGGCCGT CGCCGGCTGC GGGTGCGGTT
13201 TCGGTCCAGC CGCCGGCAGG GACAGCGCCG AACAGCTTGC TTGCATGCAG
13251 GCCGGTAGCA AAGTCCTTGA GCGTGTAGGA CGCATTGCCC TGGGGGTCCA
13301 GGTCGATCAC GGCAACCCGC AAGCCGCGCT CGAAAAAGTC GAAGGCAAGA
13351 TGCACAAGGG TCGAAGTCTT GCCGACGCCG CCTTTCTGGT TGGCCGTGAC
13401 CAAAGTTTTC ATCGTTTGGT TTCCTGTTTT TTCTTGGCGT TGGCCGTGAC
13451 CTTCCGGACG ATGTACGCCT GATGTTCCGG CAGAACCGCC GTTACCCGCG
13501 CGTACCCCTC GGGCAAGTTC TTGTCCTCGA ACGCGGCCCA CACGCGATGC
13551 ACCGCTTGCG ACACTGCGCC CCTGGTCAGT CCCAGCGACG TTGCGAACGT

13601 CGCCTGTGGC TTCCCATCGA CTAAGACGCC CCGCGCTATC TCGATGGTCT
13651 GCTGCCCCAC TTCCAGCCCC TGGATCGCCT CCTGGAACTG GCTTTCGGTA
13701 AGCCGTTTCT TCATGGATAA CACCCATAAT TTGCTCCGCG CCTTGGTTGA
13751 ACATAGCGGT GACAGCCGCC AGCACATGAG AGAAGTTTAG CTAAACATTT
13801 CTCGCACGTC AACACCTTTA GCCGCTAAAA CTCGTCCTTG GCGTAACAAA
13851 ACAAAAGCCC GGAAACCGGG CTTTCGTCTC TTGCCGCTTA TGGCTCTGCA
13901 CCCGGCTCCA TCACCAACAG GTCGCGCACG CGCTTCACTC GGTTGCGGAT
13951 CGACACTGCC AGCCCAACAA AGCCGGTTGC CGCCGCCGCC AGGATCGCGC
14001 CGATGATGCC GGCCACACCG GCCATCGCCC ACCAGGTCGC CGCCTTCCGG
14051 TTCCATTCCT GCTGGTACTG CTTCGCAATG CTGGACCTCG GCTCACCATA
14101 GGCTGACCGC TCGATGGCGT ATGCCGCTTC TCCCCTTGGC GTAAAACCCA
14151 GCGCCGCAGG CGGCATTGCC ATGCTGCCCG CCGCTTTCCC GACCACGACG
14201 CGCGCACCAG GCTTGCGGTC CAGACCTTCG GCCACGGCGA GCTGCGCAAG
14251 GACATAATCA GCCGCCGACT TGGCTCCACG CGCCTCGATC AGCTCTTGCA
14301 CTCGCGCGAA ATCCTTGGCC TCCACGGCCG CCATGAATCG CGCACGCGGC
14351 GAAGGCTCCG CAGGGCCG
```

-continued

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

pRKPLHT10Dpuf Length: 14375

1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGCG**ACTAG** TTC**CATATG**A T**AGATCT**CAC CACCACCACC ACCACCACCA
2601 CCACCAC*TAA* *TAG*GCCGGCC CTCCGTCGCG GGCGGCACCC ACGCCCGCat
2651 CGATTCCAAG GTTCAGCCAT TGAGACGGCT CCGCTTCGCG CGCAAGCGCG
2701 GGTTGGGCCG ACTGCAAGCG GAGAGGGAAG CATGGCACTG CTCAGCTTCG
2751 AGCGAAAATA TCGCGTGCCG GGGGGCACGC TGGTCGGCGG AAACCTGTTC
2801 GACTTCTGGG TCGGCCCTTT CTATGTCGGC TTCTTCGGGG TTGCGACGTT
2851 TTTCTTCGCG GCCCTGGGTA TCATTCTGAT TGCCTGGAGT GCCGTACTCC
2901 AGGGTACCTG GAACCCCCAA CTCATCTCTG TCTACCCGCC GGCCCTTGAA
2951 TATGGCCTGG GAGGTGCACC CCTCGCAAAA GGCGGGCTGT GGCAGATCAT
3001 CACGATCTGC GCCACTGGTG CCTTCGTCAG CTGGGCGCTG CGCGAAGTCG
3051 AAATCTGCCG TAAGCTGGGC ATCGGGTACC ACATCCCGTT CGCCTTCGCG
3101 TTCGCCATCC TGGCCTACCT GACGCTGGTG CTGTTCCGCC CGGTGATGAT
3151 GGGCGCCTGG GGCTATGCCT TCCCCTACGG GATCTGGACG CACCTCGACT
3201 GGGTGTCGAA CACGGGCTAC ACCTACGGCA ACTTCCACTA CAACCCTGCC
3251 CACATGATCG CCATCTCGTT CTTCTTCACG AACGCGCTGG CTCTGGCGCT
3301 GCACGGCGCC CTTGTGCTCT CCGCGGCCAA CCCCGAGAAG GGCAAGGAAA
3351 TGCGGACGCC GGATCACGAG GATACGTTCT TCCGCGATCT GGTCGGCTAC

-continued

```
3401 TCGATCGGGA CGCTCGGCAT CCACCGCCTC GGCCTGCTGC TCTCGCTGAG
3451 CGCCGTCTTC TTCAGCGCCC TCTGCATGAT CATTACCGGC ACCATCTGGT
3501 TCGATCAGTG GGTCGACTGG TGGCAATGGT GGGTGAAGCT GCCGTGGTGG
3551 GCGAACATCC CGGGAGGCAT CAATGGCTGA GTATCAGAAC ATCTTCTCCC
3601 AGGTCCAGGT CCGCGGACCG GCCGACCTGG GGATGACCGA AGACGTCAAC
3651 CTGGCCAACC GTTCGGGCGT CGGTCCCTTC TCGACCCTGC TCGGCTGGTT
3701 CGGCAACGCC CAGCTCGGCC CGATCTATCT CGGCTCGCTC GGCGTCCTGT
3751 CCCTCTTCTC GGGCCTGATG TGGTTCTTCA CCATCGGGAT CTGGTTCTGG
3801 TATCAGGCGG GCTGGAACCC GGCCGTCTTC CTGCGCGACC TGTTCTTCTT
3851 CTCGCTCGAG CCGCCGGCAC CCGAATACGG TCTGTCCTTC GCGGCTCCGC
3901 TGAAGGAAGG CGGGCTGTGG CTGATCGCGT CGTTCTTCAT GTTCGTCGCG
3951 GTCTGGTCCT GGTGGGGCCG CACCTATCTC CGCGCTCAGG CGCTGGGCAT
4001 GGGCAAGCAC ACCGCCTGGG CGTTCCTCTC GGCCATCTGG CTGTGGATGG
4051 TGCTGGGCTT CATCCGTCCG ATCCTCATGG GGTCCTGGTC GGAAGCGGTT
4101 CCCTACGGCA TCTTCTCGCA CCTCGACTGG ACGAACAACT TCTCGCTCGT
4151 CCACGGCAAC CTGTTCTACA ACCCCTTCCA CGGTCTCTCG ATCGCCTTCC
4201 TCTACGGGTC GGCCCTGCTC TTCGCGATGC ACGGTGCGAC CATCCTCGCG
4251 GTCTCCCGCT TCGGCGGCGA GCGCGAGCTG GAGCAGATCG CCGACCGCGG
4301 GACGGCAGCG GAGCGGGCCG CCCTCTTCTG GCGCTGGACC ATGGGTTTCA
4351 ACGCCACGAT GGAAGGCATC CACCGCTGGG CCATCTGGAT GGCGGTCCTC
4401 GTGACCCTCA CCGGCGGCAT CGGgATCCTG CTCTCGGGCA CGGTCGTGGA
4451 CAACTGGTAC GTCTGGGGCC AGAACCACGG CATGGCGCCG CTGAACTGAG
4501 GAGCGATCAC AATGGCTGAC AAGACCATCT TCAACGATCA CCTCAACACC
4551 AATCCGAAGA CCAACCTTCG CCTCTGGGTC GCTTTCCAGA TGATGAAGGG
4601 TGCGGGCTGG GCTGGCGGCG TGTTCTTCGG GACGCTCCTT CTCATCGGGT
4651 TCTTCCGGGT GGTCGGGCGG ATGCTTCCGA TCCAGGAGAA CCAGGCTCCG
4701 GCGCCGAACA TCACCGGCGC TCTGGAGACC GGGATCGAGC TGATCAAGCA
4751 TCTCGTCTGA GACAAGTCTC GGGGCAGGGC GGCGCGAGGC CGCCCGCTCC
4801 TCCAAGTCCG GGCCATATCG CCGGCCCGGG TCCGGGGCGA CACCACAGCC
4851 CGGTTCCCTT CCTGTTGGCG ACAGGGACCT GGTGCCGTGT GGAAGACCGC
4901 ACGGCACCCT TTTGACATTC ACGGGAGGCT CTGATGACCA ATCCCACCCC
4951 GCGACCCGAA ACCCCGCTTT TGGATCGCGT CTGCTGCCCG GCCGACATGA
5001 AGGCGCTGAG TGACGCCGAA CTGGAGCGGC TGGCCGACGA AGTGCGTTCC
5051 GAGGTCAGTG ATAGGGGTAG TTTCTTATTT TAGGCAGTTT ATATGAAATT

5101 AAGACATGCA GATGTCACAG TGGATATTGA ACTGGTCTCG AAAGCTCAAT
5151 ATCCCCCAAA GCACAAGCAC AAACTTCGAC ATCATGCAGA AGCGTTTCCC
5201 GAAccgcgtc ttcgacgtgg gcatcgccga gcagcatgcc gtgaccttcg
5251 cggccggcct cgCcgGGGcc ggGatgaagc ccttctgcgc gatctattcc
5301 tcgttcctgc aacggggtta cgaccagatc gcccatgacg tggcgctgca
5351 gaaccttccc gtccgcttcg tgatcgaccg ggcggggctc gtgggggccg
5401 atggcgcgac ccatgcgggg gccTTCGACG TTGGCTTCAT CACTTCGCTG
5451 CCCAACATGA CCGTGATGGC CGCGGCCGAC GAGGCCGAGC TCATCCACAT
5501 GATCgcCACC GCCGTGGCCT TCGACGAGGG CCCCATCGCC TTcCGCTTCC
5551 CGCGGGGCGA GGGGGTGGGC GTCGAGATGC CCGAGCGCGG GACGGTGCTG
5601 GAGCCCGgCC GGGGCCGCGT GGTGCGCGAA GGGACggatg tcgcgatcct
5651 ctccttCGGC GCGcATCTGC AcGAGgccTT GcAGGCggcG AAACTTCTCG
5701 AGGccGAGGG GGTGAGCGTG ACCGTGGCCG ACgcccgCTT ctCgCgCCCG
5751 CTCgAcACGG GGCTCATGGA CCAGCTCGTg cGCCATCACG CGGCGCTGGT
5801 AACGGTGGAG CAGGGGGCcA TGGGCGGCTT CGGCGCCCAT GTCATGCACT
5851 ATCTCGCCAA TTCCGGCGGC TTCGACGGGG GCCTCGCGCT CCGGGTCATG
5901 ACGCTGCCCG ACCGCTTCAT CGAGCAGGCG AGCCCCGAGg ACATGTATGC
5951 CGATGCGGGG CTGCGGGCCG AGGATATCAA GCTTGGCGTA ATCATGGTCA
6001 TAGCTGTTTC CTGTGTGAAA TTGTTATCCG CTCACAATTC CACACAACAT
6051 ACGAGCCGGA AGCATAAAGT GTAAAGCCTG GGGTGCCTAA TGAGTGAGCT
6101 AACTCACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC
6151 CTGTCGTGCC AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG
6201 TTTGCGTATT GGGCGCTCGG TCTTGCCTTG CTCGTCGGTG ATGTACTTCA
6251 CCAGCTCCGC GAAGTCGCTC TTCTTGATGG AGCGCATGGG GACGTGCTTG
6301 GCAATCACGC GCACCCCCCG GCCGTTTTAG CGGCTAAAAA AGTCATGGCT
6351 CTGCCCTCGG GCGGACCACG CCCATCATGA CCTTGCCAAG CTCGTCCTGC
6401 TTCTCTTCGA TCTTCGCCAG CAGGGCGAGG ATCGTGGCAT CACCGAACCG
6451 CGCCGTGCGC GGGTCGTCGG TGAGCCAGAG TTTCAGCAGG CCGCCCAGGC
6501 GGCCCAGGTC GCCATTGATG CGGGCCAGCT CGCGGACGTG CTCATAGTCC
6551 ACGACGCCCG TGATTTTGTA GCCCTGGCCG ACGGCCAGCA GGTAGGCCGA
6601 CAGGCTCATG CCGGCCGCCG CCGCCTTTTC CTCAATCGCT CTTCGTTCGT
6651 CTGGAAGGCA GTACACCTTG ATAGGTGGGC TGCCCTTCCT GGTTGGCTTG
6701 GTTTCATCAG CCATCCGCTT GCCCTCATCT GTTACGCCGG CGGTAGCCGG
6751 CCAGCCTCGC AGAGCAGGAT TCCCGTTGAG CACCGCCAGG TGCGAATAAG
```

-continued

```
 6801 GGACAGTGAA GAAGGAACAC CCGCTCGCGG GTGGGCCTAC TTCACCTATC
 6851 CTGCCCGGCT GACGCCGTTG GATACACCAA GGAAAGTCTA CACGAACCCT
 6901 TTGGCAAAAT CCTGTATATC GTGCGAAAAA GGATGGATAT ACCGAAAAAA
 6951 TCGCTATAAT GACCCCGAAG CAGGGTTATG CAGCGGAAAA GCGCCACGCT
 7001 TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
 7051 CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT
 7101 AGTCCTGTCG 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 CGGCCGTGAG
 7401 GCTTGGACGC TAGGGCAGGG CATGAAAAAG CCCGTAGCGG GCTGCTACGG
 7451 GCGTCTGACG CGGTGGAAAG GGGGAGGGGA TGTTGTCTAC ATGGCTCTGC
 7501 TGTAGTGAGT GGGTTGCGCT CCGGCAGCGG TCCTGATCAA TCGTCACCCT
 7551 TTCTCGGTCC TTCAACGTTC CTGACAACGA GCCTCCTTTT CGCCAATCCA
 7601 TCGACAATCA CCGCGAGTCC CTGCTCGAAC GCTGCGTCCG GACCGGCTTC
 7651 GTCGAAGGCG TCTATCGCGG CCCGCAACAG CGGCGAGAGC GGAGCCTGTT
 7701 CAACGGTGCC GCCGCGCTCG CCGGCATCGC TGTCGCCGGC CTGCTCCTCA
 7751 AGCACGGCCC CAACAGTGAA GTAGCTGATT GTCATCAGCG CATTGACGGC
 7801 GTCCCCGGCC GAAAAAACCG CCTCGCAGAG GAAGCGAAGC TGCGCGTCGG
 7851 CCGTTTCCAT CTGCGGTGCG CCCGGTCGCG TGCCGGCATG CATGCGCGCG
 7901 CCATCGCGGT AGGCGAGCAG CGCCTGCCTG AAGCTGCGGG CATTCCCGAT
 7951 CAGAAATGAG CGCCAGTCGT CGTCGGCTCT CGGCACCGAA TGCGTATGAT
 8001 TCTCCGCCAG CATGGCTTCG GCCAGTGCGT CGAGCAGCGC CCGCTTGTTC
 8051 CTGAAGTGCC AGTAAAGCGC CGGCTGCTGA ACCCCCAACC GTTCCGCCAG
 8101 TTTGCGTGTC GTCAGACCGT CTACGCCGAC CTCGTTCAAC AGGTCCAGGG
 8151 CGGCACGGAT CACTGTATTC GGCTGCAACT TTGTCATGCT TGACACTTTA
 8201 TCACTGATAA ACATAATATG TCCACCAACT TATCAGTGAT AAAGAATCCG
 8251 CGCGTTCAAT CGGACCAGCG GAGGCTGGTC CGGAGGCCAG ACaTGAAACC
 8301 CAACATACCC CTGATCGTAA TTCTGAGCAC TGTCGCGCTC GACGCTGTCG
 8351 GCATCGGCCT GATTATGCCG GTGCTGCCGG GCCTCCTGCG CGATCTGGTT
 8401 CACTCGAACG ACGTCACCGC CCACTATGGC ATTCTGCTGG CGCTGTATGC
 8451 GTTGGTGCAA TTTGCCTGCG CACCTGTGCT GGGCGCGCTG TCGGATCGTT

 8501 TCGGGCGGCG GCCAATCTTG CTCGTCTCGC TGGCCGGCGC CACTGTCGAC
 8551 TACGCCATCA TGGCGACAGC GCCTTTCCTT TGGGTTCTCT ATATCGGGCG
 8601 GATCGTGGCC GGCATCACCG GGGCGACTGG GGCGGTAGCC GGCGCTTATA
 8651 TTGCCGATAT CACTGATGGC GATGAGCGCG CGCGGCACTT CGGCTTCATG
 8701 AGCGCCTGTT TCGGGTTCGG GATGGTCGCG GGACCTGTGC TCGGTGGGCT
 8751 GATGGGCGGT TTCTCCCCCC ACGCTCCGTT CTTCGCCGCG GCAGCCTTGA
 8801 ACGGCCTCAA TTTCCTGACG GGCTGTTTCC TTTTGCCGGA GTCGCACAAA
 8851 GGCGAACGCC GGCCGTTACG CCGGGAGGCT CTCAACCCGC TCGCTTCGTT
 8901 CCGGTGGGCC CGGGGCATGA CCGTCGTCGC CGCCCTGATG GCGGTCTTCT
 8951 TCATCATGCA ACTTGTCGGA CAGGTGCCGG CCGCGCTTTG GGTCATTTTC
 9001 GGCGAGGATC GCTTTCACTG GGACGCGACC ACGATCGGCA TTTCGCTTGC
 9051 CGCATTTGGC ATTCTGCATT CACTCGCCCA GGCAATGATC ACCGGCCCTG
 9101 TAGCCGCCCG GCTCGGCGAA AGGCGGGCAC TCATGCTCGG AATGATTGCC
 9151 GACGGCACAG GCTACATCCT GCTTGCCTTC GCGACACGGG GATGGATGGC
 9201 GTTCCCGATC ATGGTCCTGC TTGCTTCGGG TGGCATCGGA ATGCCGGCGC
 9251 TGCAAGCAAT GTTGTCCAGG CAGGTGGATG AGGAACGTCA GGGGCAGCTG
 9301 CAAGGCTCAC TGGCGGCGCT CACCAGCCTG ACCTCGATCG TCGGACCCCT
 9351 CCTCTTCACG GCGATCTATG CGGCTTCTAT AACAACGTGG AACGGGTGGG
 9401 CATGGATTGC AGGCGCTGCC CTCTACTTGC TCTGCCTGCC GGCGCTGCGT
 9451 CGCGGGCTTT GGAGCGGCGC AGGGCAACGA GCCGATCGCT GATCGTGGAA
 9501 ACGATAGGCC TATGCCATGC GGGTCAAGGC GACTTCCGGC AAGCTATACG
 9551 CGCCCTAGGA GTGCGGTTGG AACGTTGGCC CAGCCAGATA CTCCCGATCA
 9601 CGAGCAGGAC GCCGATGATT TGAAGCGCAC TCAGCGTCTG ATCCAAGAAC
 9651 AACCATCCTA GCAACACGGC GGTCCCCGGG CTGAGAAAGC CCAGTAAGGA
 9701 AACAACTGTA GGTTCGAGTC GCGAGATCCC CCGGAACCAA AGGAAGTAGG
 9751 TTAAACCCGC TCCGATCAGG CCGAGCCACG CCAGGCCGAG AACATTGGTT
 9801 CCTGTAGGCA TCGGGATTGG CGGATCAAAC ACTAAAGCTA CTGGAACGAG
 9851 CAGAAGTCCT CCGGCCGCCA GTTGCCAGGC GGTAAAGGTG AGCAGAGGCA
 9901 CGGGAGGTTG CCACTTGCGG GTCAGCACGG TTCCGAACGC CATGGAAACC
 9951 GCCCCCGCCA GGCCCGCTGC GACGCCGACA GGATCTAGCG CTGCGTTTGG
10001 TGTCAACACC AACAGCGCCA CGCCCGCAGT TCCGCAAATA GCCCCCAGGA
10051 CCGCCATCAA TCGTATCGGG CTACCTAGCA GAGCGGCAGA GATGAACACG
10101 ACCATCAGCG GCTGCACAGC GCCTACCGTC GCCGCGACCC CGCCCGGCAG
10151 GCGGTAGACC GAAATAAACA ACAAGCTCCA GAATAGCGAA ATATTAAGTG
```

-continued

```
10201 CGCCGAGGAT GAAGATGCGC ATCCACCAGA TTCCCGTTGG AATCTGTCGG
10251 ACGATCATCA CGAGCAATAA ACCCGCCGGC AACGCCCGCA GCAGCATACC
10301 GGCGACCCCT CGGCCTCGCT GTTCGGGCTC CACGAAAACG CCGGACAGAT
10351 GCGCCTTGTG AGCGTCCTTG GGGCCGTCCT CCTGTTTGAA GACCGACAGC
10401 CCAATGATCT CGCCGTCGAT GTAGGCGCCG AATGCCACGG CATCTCGCAA
10451 CCGTTCAGCG AACGCCTCCA TGGGCTTTTT CTCCTCGTGC TCGTAAACGG
10501 ACCCGAACAT CTCTGGAGCT TTCTTCAGGG CCGACAATCG GATCTCGCGG
10551 AAATCCTGCA CGTCGGCCGC TCCAAGCCGT CGAATCTGAG CCTTAATCAC
10601 AATTGTCAAT TTTAATCCTC TGTTTATCGG CAGTTCGTAG AGCGCGCCGT
10651 GCGTCCCGAG CGATACTGAG CGAAGCAAGT GCGTCGAGCA GTGCCCGCTT
10701 GTTCCTGAAA TGCCAGTAAA GCGCTGGCTG CTGAACCCCC AGCCGGAACT
10751 GACCCCACAA GGCCCTAGCG TTTGCAATGC ACCAGGTCAT CATTGACCCA
10801 GGCGTGTTCC ACCAGGCCGC TGCCTCGCAA CTCTTCGCAG GCTTCGCCGA
10851 CCTGCTCGCG CCACTTCTTC ACGCGGGTGG AATCCGATCC GCACATGAGG
10901 CGGAAGGTTT CCAGCTTGAG CGGGTACGGC TCCCGGTGCG AGCTGAAATA
10951 GTCGAACATC CGTCGGGCCG TCGGCGACAG CTTGCGGTAC TTCTCCCATA
11001 TGAATTTCGT GTAGTGGTCG CCAGCAAACA GCACGACGAT TTCCTCGTCG
11051 ATCAGGACCT GGCAACGGGA CGTTTTCTTG CCACGGTCCA GGACGCGGAA
11101 GCGGTGCAGC AGCGACACCG ATTCCAGGTG CCCAACGCGG TCGGACGTGA
11151 AGCCCATCGC CGTCGCCTGT AGGCGCGACA GGCATTCCTC GGCCTTCGTG
11201 TAATACCGGC CATTGATCGA CCAGCCCAGG TCCTGGCAAA GCTCGTAGAA
11251 CGTGAAGGTG ATCGGCTCGC CGATAGGGGT GCGCTTCGCG TACTCCAACA
11301 CCTGCTGCCA CACCAGTTCG TCATCGTCGG CCCGCAGCTC GACGCCGGTG
11351 TAGGTGATCT TCACGTCCTT GTTGACGTGG AAAATGACCT TGTTTTGCAG
11401 CGCCTCGCGC GGGATTTTCT TGTTGCGCGT GGTGAACAGG GCAGAGCGGG
11451 CCGTGTCGTT TGGCATCGCT CGCATCGTGT CCGGCCACGG CGCAATATCG
11501 AACAAGGAAA GCTGCATTTC CTTGATCTGC TGCTTCGTGT GTTTCAGCAA
11551 CGCGGCCTGC TTGGCCTCGC TGACCTGTTT TGCCAGGTCC TCGCCGGCGG
11601 TTTTTCGCTT CTTGGTCGTC ATAGTTCCTC GCGTGTCGAT GGTCATCGAC
11651 TTCGCCAAAC CTGCCGCCTC CTGTTCGAGA CGACGCGAAC GCTCCACGGC
11701 GGCCGATGGC GCGGGCAGGG CAGGGGGAGC CAGTTGCACG CTGTCGCGCT
11751 CGATCTTGGC CGTAGCTTGC TGGACCATCG AGCCGACGGA CTGGAAGGTT
11801 TCGCGGGGCG CACGCATGAC GGTGCGGCTT GCGATGGTTT CGGCATCCTC
11851 GGCGGAAAAC CCCGCGTCGA TCAGTTCTTG CCTGTATGCC TTCCGGTCAA

11901 ACGTCCGATT CATTCACCCT CCTTGCGGGA TTGCCCCGAC TCACGCCGGG
11951 GCAATGTGCC CTTATTCCTG ATTTGACCCG CCTGGTGCCT TGGTGTCCAG
12001 ATAATCCACC TTATCGGCAA TGAAGTCGGT CCCGTAGACC GTCTGGCCGT
12051 CCTTCTCGTA CTTGGTATTC CGAATCTTGC CCTGCACGAA TACCAGCGAC
12101 CCCTTGCCCA AATACTTGCC GTGGGCCTCG GCCTGAGAGC CAAAACACTT
12151 GATGCGGAAG AAGTCGGTGC GCTCCTGCTT GTCGCCGGTC GTGGCCGCGC
12201 CAACCTTTGC GATCCGCAAG CGCGCGGTCG CCATCTTCAC GCTGGAACAG
12251 TACGTCGAGG CGGGCATCAT GACCCGCGAG CAATACGAGG TCATTAAAAG
12301 CGCCGTGATT GATGATATAG CGGCCCGGCT GCTCCTGGTT CTCGCGCACC
12351 GAAATGGGTG ACTTCACCCC GCGCTCTTTG ATCGTGGCAC CGATTTCCGC
12401 GATGCTCTCC GGGGAAAAGC CGGGTTTGTC GGCCGTCCGC GGCTGATGCG
12451 GATCTTCGTC GATCAGGTCC AGGTCCAGCT CGATAGGGCC GGAACCGCCC
12501 TGAGACGCCG CAGGAGCGTC CAGGAGGCTC GACAGGTCGC CGATGCTATC
12551 CAACCCCAGG CCGGACGGCT GCGCCGCGCC TGCGGCTTCC TGAGCGGCCG
12601 CAGCGGTGTT TTTCTTGGTG GTCTTGGCTT GAGCCGCAGT CATTGGGAAA
12651 TCTCCATCTT CGTGAACACG TAATCAGCCA GGGCGCGAAC CTCTTTCGAT
12701 GCCTTGCGCG CGGCCGTTTT CTTGATCTTC CAGACCGGCA CACCGGATGC
12751 GAGGGCATCG GCGATGCTGC TGCGCAGGCC AACGGTGGCC GGAATCATCA
12801 TCTTGGGGTA CGCGGCCAGC AGCTCGGCTT GGTGGCGCGC GTGGCGCGGA
12851 TTCCGCGCAT CGACCTTGCT GGGCACCATG CCAAGGAATT GCAGCTTGGC
12901 GTTCTTCTGG CGCACGTTCG CAATGGTCGT GACCATCTTC TTGATGCCCT
12951 GGATGCTGTA CGCCTCAAGC TCGATGGGGG ACAGCACATA GTCGGCCGCG
13001 AAGAGGGCGG CCGCCAGGCC GACGCCAAGG GTCGGGGCCG TGTCGATCAG
13051 GCACACGTCG AAGCCTTGGT TCGCCAGGGC CTTGATGTTC GCCCCGAACA
13101 GCTCGCGGGC GTCGTCCAGC GACAGCCGTT CGGCGTTCGC CAGTACCGGG
13151 TTGGACTCGA TGAGGGCGAG GCGCGCGGCC TGGCCGTCGC CGGCTGCGGG
13201 TGCGGTTTCG GTCCAGCCGC CGGCAGGGAC AGCGCCGAAC AGCTTGCTTG
13251 CATGCAGGCC GGTAGCAAAG TCCTTGAGCG TGTAGGACGC ATTGCCCTGG
13301 GGGTCCAGGT CGATCACGGC AACCCGCAAG CCGCGCTCGA AAAAGTCGAA
13351 GGCAAGATGC ACAAGGGTCG AAGTCTTGCC GACGCCGCCT TTCTGGTTGG
13401 CCGTGACCAA AGTTTTCATC GTTTGGTTTC CTGTTTTTTC TTGGCGTCCG
13451 CTTCCCACTT CCGGACGATG TACGCCTGAT GTTCCGGCAG AACCGCCGTT
13501 ACCCGCGCGT ACCCCTCGGG CAAGTTCTTG TCCTCGAACG CGGCCCACAC
13551 GCGATGCACC GCTTGCGACA CTGCGCCCCT GGTCAGTCCC AGCGACGTTG

13601 CGAACGTCGC CTGTGGCTTC CCATCGACTA AGACGCCCCG CGCTATCTCG
13651 ATGGTCTGCT GCCCCACTTC CAGCCCCTGG ATCGCCTCCT GGAACTGGCT
13701 TTCGGTAAGC CGTTTCTTCA TGGATAACAC CCATAATTTG CTCCGCGCCT
13751 TGGTTGAACA TAGCGGTGAC AGCCGCCAGC ACATGAGAGA AGTTTAGCTA
13801 AACATTTCTC GCACGTCAAC ACCTTTAGCC GCTAAAACTC GTCCTTGGCG
13851 TAACAAAACA AAAGCCCGGA AACCGGGCTT TCGTCTCTTG CCGCTTATGG
13901 CTCTGCACCC GGCTCCATCA CCAACAGGTC GCGCACGCGC TTCACTCGGT
13951 TGCGGATCGA CACTGCCAGC CCAACAAAGC CGGTTGCCGC CGCCGCCAGG
14001 ATCGCGCCGA TGATGCCGGC CACACCGGCC ATCGCCCACC AGGTCGCCGC
14051 CTTCCGGTTC CATTCCTGCT GGTACTGCTT CGCAATGCTG GACCTCGGTA
14101 CACCATAGGC TGACCGCTCG ATGGCGTATG CCGCTTCTCC CCTTGGCGTA
14151 AAACCCAGCG CCGCAGGCGG CATTGCCATG CTGCCCGCCG CTTTCCCGAC
14201 CACGACGCGC GCACCAGGCT TGCGGTCCAG ACCTTCGGCC ACGGCGAGCT
14251 GCGCAAGGAC ATAATCAGCC GCCGACTTGG CTCCACGCGC CTCGATCAGC
14301 TCTTGCACTC GCGCGAAATC CTTGGCCTCC ACGGCCGCCA TGAATCGCGC
14351 ACGCGGCGAA GGCTCCGCAG GGCCG
```

-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 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCgTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGCG**ACTAG** TTC**CATATG**A **TAGATCT**CAC CACCACCACC ACCACCACCA
2601 CCACCACCAC CACCAC*TAAT* *AG*GCCGGCCC TCCGTCGCGG GCGGCACCCA
2651 CGCCCGCatC GATTCCAAGG TTCAGCCATT GAGACGGCTC CGCTTCGCGC
2701 GCAAGCGCGG GTTGGGCCGA CTGCAAGCGG AGAGGGAAGC ATGGCACTGC
2751 TCAGCTTCGA GCGAAAATAT CGCGTGCCGG GGGGCACGCT GGTCGGCGGA
2801 AACCTGTTCG ACTTCTGGGT CGGCCCTTTC TATGTCGGCT TCTTCGGGGT
2851 TGCGACGTTT TTCTTCGCGG CCCTGGGTAT CATTCTGATT GCCTGGAGTG
2901 CCGTACTCCA GGGTACCTGG AACCCCCAAC TCATCTCTGT CTACCCGCCG
2951 GCCCTTGAAT ATGGCCTGGG AGGTGCACCC CTCGCAAAAG GCGGGCTGTG
3001 GCAGATCATC ACGATCTGCG CCACTGGTGC CTTCGTCAGC TGGGCGCTGC
3051 GCGAAGTCGA AATCTGCCGT AAGCTGGGCA TCGGGTACCA CATCCCGTTC
3101 GCCTTCGCGT TCGCCATCCT GGCCTACCTG ACGCTGGTGC TGTTCCGCCC
3151 GGTGATGATG GGCGCCTGGG GCTATGCCTT CCCCTACGGG ATCTGGACGC
3201 ACCTCGACTG GGTGTCGAAC ACGGGCTACA CCTACGGCAA CTTCCACTAC
3251 AACCCTGCCC ACATGATCGC CATCTCGTTC TTCTTCACGA ACGCGCTGGC
3301 TCTGGCGCTG CACGGCGCCC TTGTGCTCTC CGCGGCCAAC CCCGAGAAGG
3351 GCAAGGAAAT GCGGACGCCG GATCACGAGG ATACGTTCTT CCGCGATCTG

-continued

```
3401 GTCGGCTACT CGATCGGGAC GCTCGGCATC CACCGCCTCG GCCTGCTGCT
3451 CTCGCTGAGC GCCGTCTTCT TCAGCGCCCT CTGCATGATC ATTACCGGCA
3501 CCATCTGGTT CGATCAGTGG GTCGACTGGT GGCAATGGTG GGTGAAGCTG
3551 CCGTGGTGGG CGAACATCCC GGGAGGCATC AATGGCTGAG TATCAGAACA
3601 TCTTCTCCCA GGTCCAGGTC CGCGGACCGG CCGACCTGGG GATGACCGAA
3651 GACGTCAACC TGGCCAACCG TTCGGGCGTC GGTCCCTTCT CGACCCTGCT
3701 CGGCTGGTTC GGCAACGCCC AGCTCGGCCC GATCTATCTC GGCTCGCTCG
3751 GCGTCCTGTC CCTCTTCTCG GGCCTGATGT GGTTCTTCAC CATCGGGATC
3801 TGGTTCTGGT ATCAGGCGGG CTGGAACCCG GCCGTCTTCC TGCGCGACCT
3851 GTTCTTCTTC TCGCTCGAGC CGCCGGCACC CGAATACGGT CTGTCCTTCG
3901 CGGCTCCGCT GAAGGAAGGC GGGCTGTGGC TGATCGCGTC GTTCTTCATG
3951 TTCGTCGCGG TCTGGTCCTG GTGGGGCCGC ACCTATCTCC GCGCTCAGGC
4001 GCTGGGCATG GGCAAGCACA CCGCCTGGGC GTTCCTCTCG GCCATCTGGC
4051 TGTGGATGGT GCTGGGCTTC ATCCGTCCGA TCCTCATGGG GTCCTGGTCG
4101 GAAGCGGTTC CCTACGGCAT CTTCTCGCAC CTCGACTGGA CGAACAACTT
4151 CTCGCTCGTC CACGGCAACC TGTTCTACAA CCCCTTCCAC GGTCTCTCGA
4201 TCGCCTTCCT CTACGGGTCG GCCCTGCTCT TCGCGATGCA CGGTGCGACC
4251 ATCCTCGCGG TCTCCCGCTT CGGCGGCGAG CGCGAGCTGG AGCAGATCGC
4301 CGACCGCGGG ACGGCAGCGG AGCGGGCCGC CCTCTTCTGG CGCTGGACCA
4351 TGGGTTTCAA CGCCACGATG GAAGGCATCC ACCGCTGGGC CATCTGGATG
4401 GCGGTCCTCG TGACCCTCAC CGGCGGCATC GGgATCCTGC TCTCGGGCAC
4451 GGTCGTGGAC AACTGGTACG TCTGGGGCCA GAACCACGGC ATGGCGCCGC
4501 TGAACTGAGG AGCGATCACA ATGGCTGACA AGACCATCTT CAACGATCAC
4551 CTCAACACCA ATCCGAAGAC CAACCTTCGC CTCTGGGTCG CTTTCCAGAT
4601 GATGAAGGGT GCGGGCTGGG CTGGCGGCGT GTTCTTCGGG ACGCTCCTTC
4651 TCATCGGGTT CTTCCGGGTG GTCGGGCGGA TGCTTCCGAT CCAGGAGAAC
4701 CAGGCTCCGG CGCCGAACAT CACCGGCGCT CTGGAGACCG GGATCGAGCT
4751 GATCAAGCAT CTCGTCTGAG ACAAGTCTCG GGGCAGGGCG GCGCGAGGCC
4801 GCCCGCTCCT CCAAGTCCGG GCCATATCGC CGGCCCGGGT CCGGGGCGAC
4851 ACCACAGCCC GGTTCCCTTC CTGTTGGCGA CAGGGACCTG GTGCCGTGTG
4901 GAAGACCGCA CGGCACCCTT TTGACATTCA CGGGAGGCTC TGATGACCAA
4951 TCCCACCCCG CGACCCGAAA CCCCGCTTTT GGATCGCGTC TGCTGCCCGG
5001 CCGACATGAA GGCGCTGAGT GACGCCGAAC TGGAGCGGCT GGCCGACGAA
5051 GTGCGTTCCG AGGTCAGTGA TAGGGGTAGT TTCTTATTTT AGGCAGTTTA

5101 TATGAAATTA AGACATGCAG ATGTCACAGT GGATATTGAA CTGGTCTCGA
5151 AAGCTCAATA TCCCCCAAAG CACAAGCACA AACTTCGACA TCATGCAGAA
5201 GCGTTTCCCG AAccgcgtct tcgacgtggg catcgccgag cagcatgccg
5251 tgaccttcgc ggccggcctc gCcgGGGccg gGatgaagcc cttctgcgcg
5301 atctattcct cgttcctgca acggggttac gaccagatcg cccatgacgt
5351 ggcgctgcag aaccttcccg tccgcttcgt gatcgaccgg gcggggctcg
5401 tgggggccga tggcgcgacc catgcggggg ccTTCGACGT TGGCTTCATC
5451 ACTTCGCTGC CCAACATGAC CGTGATGGCC GCGGCCGACG AGGCCGAGCT
5501 CATCCACATG ATCgcCACCG CCGTGGCCTT CGACGAGGGC CCCATCGCCT
5551 TcCGCTTCCC GCGGGGCGAG GGGGTGGGCG TCGAGATGCC CGAGCGCGGG
5601 ACGGTGCTGG AGCCCGgCCG GGGCCGCGTG GTGCGCGAAG GGACggatgt
5651 cgcgatcctc tccttCGGCG CGcATCTGCA cGAGgccTTG cAGGCggcGA
5701 AACTTCTCGA GGccGAGGGG GTGAGCGTGA CCGTGGCCGA CgcccgCTTc
5751 tCgCgCCCGC TCgAcACGGG GCTCATCGAC CAGCTCGTgc GCCATCACGC
5801 GGCGCTGGTA ACGGTGGAGC AGGGGGCcAT GGGCGGCTTC GGCGCCCATG
5851 TCATGCACTA TCTCGCCAAT TCCGGCGGCT TCGACGGGGG CCTCGCGCTC
5901 CGGGTCATGA CGCTGCCCGA CCGCTTCATC GAGCAGGCGA GCCCCGAGgA
5951 CATGTATGCC GATGCGGGGC TGCGGGCCGA GGATATCAAG CTTGGCGTAA
6001 TCATGGTCAT AGCTGTTTCC TGTGTGAAAT TGTTATCCGC TCACAATTCC
6051 ACACAACATA CGAGCCGGAA GCATAAAGTG TAAAGCCTGG GGTGCCTAAT
6101 GAGTGAGCTA ACTCACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCCAG
6151 TCGGGAAACC TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGCGGG
6201 GAGAGGCGGT TTGCGTATTG GGCGCTCGGT CTTGCCTTGC TCGTCGGTGA
6251 TGTACTTCAC CAGCTCCGCG AAGTCGCTCT TCTTGATGGA GCGCATGGGG
6301 ACGTGCTTGG CAATCACGCG CACCCCCCGG CCGTTTTAGC GGCTAAAAAA
6351 GTCATGGCTC TGCCCTCGGG CGGACCACGC CCATCATGAC CTTGCCAAGC
6401 TCGTCCTGCT TCTCTTCGAT CTTCGCCAGC AGGGCGAGGA TCGTGGCATC
6451 ACCGAACCGC GCCGTGCGCG GGTCGTCGGT GAGCCAGAGT TTCAGCAGGC
6501 CGCCCAGGCG GCCCAGGTCG CCATTGATGC GGGCCAGCTC GCGGACGTGC
6551 TCATAGTCCA CGACGCCCGT GATTTTGTAG CCCTGGCCGA CGGCCAGCAG
6601 GTAGGCCGAC AGGCTCATGC CGGCCGCCGC CGCCTTTTCC TCAATCGCTC
6651 TTCGTTCGTC TGGAAGGCAG TACACCTTGA TAGGTGGGCT GCCCTTCCTG
6701 GTTGGCTTGG TTTCATCAGC CATCCGCTTG CCCTCATCTG TTACGCCGGC
6751 GGTAGCCGGC CAGCCTCGCA GAGCAGGATT CCCGTTGAGC ACCGCCAGGT
```

-continued

```
 6801 GCGAATAAGG GACAGTGAAG AAGGAACACC CGCTCGCGGG TGGGCCTACT
 6851 TCACCTATCC TGCCCGGCTG ACGCCGTTGG ATACACCAAG GAAAGTCTAC
 6901 ACGAACCCTT TGGCAAAATC CTGTATATCG TGCGAAAAAG GATGGATATA
 6951 CCGAAAAAAT CGCTATAATG ACCCCGAAGC AGGGTTATGC AGCGGAAAAG
 7001 CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG GTAAGCGGCA
 7051 GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG
 7101 TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG AGCGTCGATT
 7151 TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG
 7201 CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC
 7251 TTTCCTGCGT TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCTTTGA
 7301 GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG
 7351 TGAGCGAGGA AGCGGAAGAG CGCCAGAAGG CCGCCAGAGA GGCCGAGCGC
 7401 GGCCGTGAGG CTTGGACGCT AGGGCAGGGC ATGAAAAAGC CCGTAGCGGG
 7451 CTGCTACGGG CGTCTGACGC GGTGGAAAGG GGGAGGGGAT GTTGTCTACA
 7501 TGGCTCTGCT GTAGTGAGTG GGTTGCGCTC CGGCAGCGGT CCTGATCAAT
 7551 CGTCACCCTT TCTCGGTCCT TCAACGTTCC TGACAACGAG CCTCCTTTTC
 7601 GCCAATCCAT CGACAATCAC CGCGAGTCCC TGCTCGAACG CTGCGTCCGG
 7651 ACCGGCTTCG TCGAAGGCGT CTATCGCGGC CCGCAACAGC GGCGAGAGCG
 7701 GAGCCTGTTC AACGGTGCCG CCGCGCTCGC CGGCATCGCT GTCGCCGGCC
 7751 TGCTCCTCAA GCACGGCCCC AACAGTGAAG TAGCTGATTG TCATCAGCGC
 7801 ATTGACGGCG TCCCCGGCCG AAAAACCCGC CTCGCAGAGG AAGCGAAGCT
 7851 GCGCGTCGGC CGTTTCCATC TGCGGTGCGC CCGGTCGCGT GCCGGCATGG
 7901 ATGCGCGCGC CATCGCGGTA GGCGAGCAGC GCCTGCCTGA AGCTGCGGGC
 7951 ATTCCCGATC AGAAATGAGC GCCAGTCGTC GTCGGCTCTC GGCACCGAAT
 8001 GCGTATGATT CTCCGCCAGC ATGGCTTCGG CCAGTGCGTC GAGCAGCGCC
 8051 CGCTTGTTCC TGAAGTGCCA GTAAAGCGCC GGCTGCTGAA CCCCCAACCG
 8101 TTCCGCCAGT TTGCGTGTCG TCAGACCGTC TACGCCGACC TCGTTCAACA
 8151 GGTCCAGGGC GGCACGGATC ACTGTATTCG GCTGCAACTT TGTCATGCTT
 8201 GACACTTTAT CACTGATAAA CATAATATGT CCACCAACTT ATCAGTGATA
 8251 AAGAATCCGC GCGTTCAATC GGACCAGCGG AGGCTGGTCC GGAGGCCAGA
 8301 CaTGAAACCC AACATACCCC TGATCGTAAT TCTGAGCACT GTCGCGCTCG
 8351 ACGCTGTCGG CATCGGCCTG ATTATGCCGG TGCTGCCGGG CCTCCTGCGC
 8401 GATCTGGTTC ACTCGAACGA CGTCACCGCC CACTATGGCA TTCTGCTGGC
 8451 GCTGTATGCG TTGGTGCAAT TTGCCTGCGC ACCTGTGCTG GGCGCGCTGT

 8501 CGGATCGTTT CGGGCGGCGG CCAATCTTGC TCGTCTCGCT GGCCGGCGCC
 8551 ACTGTCGACT ACGCCATCAT GGCGACAGCG CCTTTCCTTT GGGTTCTCTA
 8601 TATCGGGCGG ATCGTGGCCG GCATCACCGG GGCGACTGGG GCGGTAGCCG
 8651 GCGCTTATAT TGCCGATATC ACTGATGGCG ATGAGCGCGC GCGGCACTTC
 8701 GGCTTCATGA GCGCCTGTTT CGGGTTCGGG ATGGTCGCGG GACCTGTGCT
 8751 CGGTGGGCTG ATGGGCGGTT TCTCCCCCCA CGCTCCGTTC TTCGCCGCGG
 8801 CAGCCTTGAA CGGCCTCAAT TTCCTGACGG GCTGTTTCCT TTTGCCGGAG
 8851 TCGCACAAAG GCGAACGCCG GCCGTTACGC CGGGAGGCTC TCAACCCGCT
 8901 CGCTTCGTTC CGGTGGGCCC GGGGCATGAC CGTCGTCGCC GCCCTGATGG
 8951 CGGTCTTCTT CATCATGCAA CTTGTCGGAC AGGTGCCGGC CGCGCTTTGG
 9001 GTCATTTTCG GCGAGGATCG CTTTCACTGG GACGCGACCA CGATCGGCAT
 9051 TTCGCTTGCC GCATTTGGCA TTCTGCATTC ACTCGCCCAG GCAATGATCA
 9101 CCGGCCCTGT AGCCGCCCGG CTCGGCGAAA GGCGGGCACT CATGCTCGGA
 9151 ATGATTGCCG ACGGCACAGG CTACATCCTG CTTGCCTTCG CGACACGGGG
 9201 ATGGATGGCG TTCCCGATCA TGGTCCTGCT TGCTTCGGGT GGCATCGGAA
 9251 TGCCGGCGCT GCAAGCAATG TTGTCCAGGC AGGTGGATGA GGAACGTCAG
 9301 GGGCAGCTGC AAGGCTCACT GGCGGCGCTC ACCAGCCTGA CCTCGATCGT
 9351 CGGACCCCTC CTCTTCACGG CGATCTATGC GGCTTCTATA ACAACGTGGA
 9401 ACGGGTGGGC ATGGATTGCA GGCGCTGCCC TCTACTTGCT CTGCCTGCCG
 9451 GCGCTGCGTC GCGGGCTTTG GAGCGGCGCA GGGCAACGAG CCGATCGCTG
 9501 ATCGTGGAAA CGATAGGCCT ATGCCATGCG GGTCAAGGCG ACTTCCGGCA
 9551 AGCTATACGC GCCCTAGGAG TGCGGTTGGA ACGTTGGCCC AGCCAGATAC
 9601 TCCCGATCAC GAGCAGGACG CCGATGATTT GAAGCGCACT CAGCGTCTGA
 9651 TCCAAGAACA ACCATCCTAG CAACACGGCG GTCCCCGGGC TGAGAAAGCC
 9701 CAGTAAGGAA ACAACTGTAG GTTCGAGTCG CGAGATCCCC CGGAACCAAA
 9751 GGAAGTAGGT TAAACCCGCT CCGATCAGGC CGAGCCACGC CAGGCCGAGA
 9801 ACATTGGTTC CTGTAGGCAT CGGGATTGGC GGATCAAACA CTAAAGCTAC
 9851 TGGAACGAGC AGAAGTCCTC CGGCCGCCAG TTGCCAGGCG GTAAAGGTGA
 9901 GCAGAGGCAC GGGAGGTTGC CACTTGCGGG TCAGCACGGT TCCGAACGCC
 9951 ATGGAAACCG CCCCCGCCAG GCCCGCTGCG ACGCCGACAG GATCTAGCGC
10001 TGCGTTTGGT GTCAACACCA ACAGCGCCAC GCCCGCAGTT CCGCAAATAG
10051 CCCCCAGGAC CGCCATCAAT CGTATCGGGC TACCTAGCAG AGCGGCAGAG
10101 ATGAACACGA CCATCAGCGG CTGCACAGCG CCTACCGTCG CCGCGACCCC
10151 GCCCGGCAGG CGGTAGACCG AAATAAACAA CAAGCTCCAG AATAGCGAAA
```

-continued

```
10201 TATTAAGTGC GCCGAGGATG AAGATGCGCA TCCACCAGAT TCCCGTTGGA
10251 ATCTGTCGGA CGATCATCAC GAGCAATAAA CCCGCCGGCA ACGCCCGCAG
10301 CAGCATACCG GCGACCCCTC GGCCTCGCTG TTCGGGCTCC ACGAAAACGC
10351 CGGACAGATG CGCCTTGTGA GCGTCCTTGG GGCCGTCCTC CTGTTTGAAG
10401 ACCGACAGCC CAATGATCTC GCCGTCGATG TAGGCGCCGA ATGCCACGGC
10451 ATCTCGCAAC CGTTCAGCGA ACGCCTCCAT GGGCTTTTTC TCCTCGTGCT
10501 CGTAAACGGA CCCGAACATC TCTGGAGCTT TCTTCAGGGC CGACAATCGG
10551 ATCTCGCGGA AATCCTGCAC GTCGGCCGCT CCAAGCCGTC GAATCTGAGC
10601 CTTAATCACA ATTGTCAATT TTAATCCTCT GTTTATCGGC AGTTCGTAGA
10651 GCGCGCCGTG CGTCCCGAGC GATACTGAGC GAAGCAAGTG CGTCGAGCAG
10701 TGCCCGCTTG TTCCTGAAAT GCCAGTAAAG CGCTGGCTGC TGAACCCCCA
10751 GCCGGAACTG ACCCCACAAG GCCCTAGCGT TTGCAATGCA CCAGGTCATC
10801 ATTGACCCAG GCGTGTTCCA CCAGGCCGCT GCCTCGCAAC TCTTCGCAGG
10851 CTTCGCCGAC CTGCTCGCGC CACTTCTTCA CGCGGGTGGA ATCCGATCCG
10901 CACATGAGGC GGAAGGTTTC CAGCTTGAGC GGGTACGGCT CCCGGTGCGA
10951 GCTGAAATAG TCGAACATCC GTCGGGCCGT CGGCGACAGC TTGCGGTACT
11001 TCTCCCATAT GAATTTCGTG TAGTGGTCGC CAGCAAACAG CACGACGATT
11051 TCCTCGTCGA TCAGGACCTG GCAACGGGAC GTTTTCTTGC CACGGTCCAG
11101 GACGCGGAAG CGGTGCAGCA GCGACACCGA TTCCAGGTGC CCAACGCGGT
11151 CGGACGTGAA GCCCATCGCC GTCGCCTGTA GGCGCGACAG GCATTCCTCG
11201 GCCTTCGTGT AATACCGGCC ATTGATCGAC CAGCCCAGGT CCTGGCAAAG
11251 CTCGTAGAAC GTGAAGGTGA TCGGCTCGCC GATAGGGGTG CGCTTCGCGT
11301 ACTCCAACAC CTGCTGCCAC ACCAGTTCGT CATCGTCGGC CCGCAGCTCG
11351 ACGCCGGTGT AGGTGATCTT CACGTCCTTG TTGACGTGGA AAATGACCTT
11401 GTTTTGCAGC GCCTCGCGCG GGATTTTCTT GTTGCGCGTG GTGAACAGGG
11451 CAGAGCGGGC CGTGTCGTTT GGCATCGCTC GCATCGTGTC CGGCCACGGC
11501 GCAATATCGA ACAAGGAAAG CTGCATTTCC TTGATCTGCT GCTTCGTGTG
11551 TTTCAGCAAC GCGGCCTGCT TGGCCTCGCT GACCTGTTTT GCCAGGTCCT
11601 CGCCGGCGGT TTTTCGCTTC TTGGTCGTCA TAGTTCCTCG CGTGTCGATG
11651 GTCATCGACT TCGCCAAACC TGCCGCCTCC TGTTCGAGAC GACGCGAACG
11701 CTCCACGGCG GCCGATGGCG CGGGCAGGGC AGGGGGAGCC AGTTGCACGC
11751 TGTCGCGCTC GATCTTGGCC GTAGCTTGCT GGACCATCGA GCCGACGGAC
11801 TGGAAGGTTT CGCGGGGCGC ACGCATGACG GTGCGGCTTG CGATGGTTTC
11851 GGCATCCTCG GCGGAAAACC CCGCGTCGAT CAGTTCTTGC CTGTATGCCT

11901 TCCGGTCAAA CGTCCGATTC ATTCACCCTC CTTGCGGGAT TGCCCCGACT
11951 CACGCCGGGG CAATGTGCCC TTATTCCTGA TTTGACCCGC CTGGTGCCTT
12001 GGTGTCCAGA TAATCCACCT TATCGGCAAT GAAGTCGGTC CCGTAGACCG
12051 TCTGGCCGTC CTTCTCGTAC TTGGTATTCC GAATCTTGCC CTGCACGAAT
12101 ACCAGCGACC CCTTGCCCAA ATACTTGCCG TGGGCCTCGG CCTGAGAGCC
12151 AAAACACTTG ATGCGGAAGA AGTCGGTGCG CTCCTGCTTG TCGCCGGTCG
12201 TGGCCGCGCC AACCTTTGCG ATCCGCAAGC GCGCGGTCGC CATCTTCACG
12251 CTGGAACAGT ACGTCGAGGC GGGCATCATG ACCCGCGAGC AATACGAGGT
12301 CATTAAAAGC GCCGTGATTG ATGATATAGC GGCCCGGCTG CTCCTGGTTC
12351 TCGCGCACCG AAATGGGTGA CTTCACCCCG CGCTCTTTGA TCGTGGCACC
12401 GATTTCCGCG ATGCTCTCCG GGGAAAAGCC GGGGTTGTCG GCCGTCCGCG
12451 GCTGATGCGG ATCTTCGTCG ATCAGGTCCA GGTCCAGCTC GATAGGGCCG
12501 GAACCGCCCT GAGACGCCGC AGGAGCGTCC AGGAGGCTCG ACAGGTCGCC
12551 GATGCTATCC AACCCCAGGC CGGACGGCTG CGCCGCGCCT GCGGCTTCCT
12601 GAGCGGCCGC AGCGGTGTTT TTCTTGGTGG TCTTGGCTTG AGCCGCAGTC
12651 ATTGGGAAAT CTCCATCTTC GTGAACACGT AATCAGCCAG GGCGCGAACC
12701 TCTTTCGATG CCTTGCGCGC GGCCGTTTTC TTGATCTTCC AGACCGGCAC
12751 ACCGGATGCG AGGGCATCGG CGATGCTGCT GCGCAGGCCA ACGGTGGCCG
12801 GAATCATCAT CTTGGGGTAC GCGGCCAGCA GCTCGGCTTG GTGGCGCGCG
12851 TGGCGCGGAT TCCGCGCATC GACCTTGCTG GGCACCATGC CAAGGAATTG
12901 CAGCTTGGCG TTCTTCTGGC GCACGTTCGC AATGGTCGTG ACCATCTTCT
12951 TGATGCCCTG GATGCTGTAC GCCTCAAGCT CGATGGGGGA CAGCACATAG
13001 TCGGCCGCGA AGAGGGCGGC CGCCAGGCCG ACGCCAAGGG TCGGGGCCGT
13051 GTCGATCAGG CACACGTCGA AGCCTTGGTT CGCCAGGGCC TTGATGTTCG
13101 CCCCGAACAG CTCGCGGGCG TCGTCCAGCG ACAGCCGTTC GGCGTTCGCC
13151 AGTACCGGGT TGGACTCGAT GAGGGCGAGG CGCGCGGCCT GGCCGTCGCC
13201 GGCTGCGGGT GCGGTTTCGG TCCAGCCGCC GGCAGGGACA GCGCCGAACA
13251 GCTTGCTTGC ATGCAGGCCG GTAGCAAAGT CCTTGAGCGT GTAGGACGCA
13301 TTGCCCTGGG GGTCCAGGTC GATCACGGCA ACCCGCAAGC CGCGCTCGAA
13351 AAAGTCGAAG GCAAGATGCA CAAGGGTCGA AGTCTTGCCG ACGCCGCCTT
13401 TCTGGTTGGC CGTGACCAAA GTTTTCATCG TTTGGTTTCC TGTTTTTTCT
13451 TGGCGTCCGC TTCCCACTTC CGGACGATGT ACGCCTGATG TTCCGGCAGA
13501 ACCGCCGTTA CCCGCGCGTA CCCCTCGGGC AAGTTCTTGT CCTCGAACGC
13551 GGCCCACACG CGATGCACCG CTTGCGACAC TGCGCCCCTG GTCAGTCCCA

13601 GCGACGTTGC GAACGTCGCC TGTGGCTTCC CATCGACTAA GACGCCCCGC
13651 GCTATCTCGA TGGTCTGCTG CCCCACTTCC AGCCCCTGGA TCGCCTCCTG
13701 GAACTGGCTT TCGGTAAGCC GTTTCTTCAT GGATAACACC CATAATTTGC
13751 TCCGCGCCTT GGTTGAACAT AGCGGTGACA GCCGCCAGCA CATGAGAGAA
13801 GTTTAGCTAA ACATTTCTCG CACGTCAACA CCTTTAGCCG CTAAAACTCG
13851 TCCTTGGCGT AACAAAACAA AAGCCCCGGA ACCGGGCTTT CGTCTCTTGC
13901 CGCTTATGGC TCTGCACCCG GCTCCATCAC CAACAGGTCG CGCACGCGCT
13951 TCACTCGGTT GCGGATCGAC ACTGCCAGCC CAACAAAGCC GGTTGCCGCC
14001 GCCGCCAGGA TCGCGCCGAT GATGCCGGCC ACACCGGCCA TCGCCCACCA
14051 GGTCGCCGCC TTCCGGTTCC ATTCCTGCTG GTACTGCTTC GCAATGCTGG
14101 ACCTCGGCTC ACCATAGGCT GACCGCTCGA TGGCGTATGC CGCTTCTCCC
14151 CTTGGCGTAA AACCCAGCGC CGCAGGCGGC ATTGCCATGC TGCCCGCCGC
14201 TTTCCCGACC ACGACGCGCG CACCAGGCTT GCGGTCCAGA CCTTCGGCCA
14251 CGGCGAGCTG CGCAAGGACA TAATCAGCCG CCGACTTGGC TCCACGCGCC
14301 TCGATCAGCT CTTGCACTCG CGCGAAATCC TTGGCCTCCA CGGCCGCCAT
14351 GAATCGCGCA CGCGGCGAAG GCTCCGCAGG GCCG
```

-continued

pRKMAHT1Dpuf.seq

1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCGTGGCCC GCAAGAGGGG

1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGCGACTAG GGGAGGATAG TAATGCTCGT CAAGACGCAC ATCACGAAGA
2601 TTGGGGTCAC GCTGTTCGCC GTGGCTCTGT TCTACGGGTT CATCTACATG
2651 CTTTCCAACT CGCTCTTTGC GACGCGGCCG GCCACCGCCG TTGCGGTCGG
2701 TGCGGACGGG AAAGCCCTG**A CTAGT**TCCAT ATGAT**AGATC T**CACCACCAC
2751 CACCACCACC AC*TAATAG*GC CGGCCCTCCG TCGCGGGCGG CACCCACGCC
2801 CGCATCGATT CCAAGGTTCA GCCATTGAGA CGGCTCCGCT TCGCGCGCAA
2851 GCGCGGGTTG GGCCGACTGC AAGCGGAGAG GGAAGCATGG CACTGCTCAG
2901 CTTCGAGCGA AAATATCGCG TGCCGGGGGG CACGCTGGTC GGCGGAAACC
2951 TGTTCGACTT CTGGGTCGGC CCTTTCTATG TCGGCTTCTT CGGGGTTGCG
3001 ACGTTTTTCT TCGCGGCCCT GGGTATCATT CTGATTGCCT GGAGTGCCGT
3051 ACTCCAGGGT ACCTGGAACC CCCAACTCAT CTCTGTCTAC CCGCCGGCCC
3101 TTGAATATGG CCTGGGAGGT GCACCCCTCG CAAAAGGCGG GCTGTGGCAG
3151 ATCATCACGA TCTGCGCCAC TGGTGCCTTC GTCAGCTGGGCGCTGCGCGA
3201 AGTCGAAATC TGCCGTAAGC TGGGCATCGG GTACCACATC CCGTTCGCCT
3251 TCGCGTTCGC CATCCTGGCC TACCTGACGC TGGTGCTGTT CCGCCCGGTG
3301 ATGATGGGCG CCTGGGGCTA TGCCTTCCCC TACGGGATCT GGACGCACCT
3351 CGACTGGGTG TCGAACACGG GCTACACCTA CGGCAACTTC CACTACAACC

-continued

```
3401 CTGCCCACAT GATCGCCATC TCGTTCTTCT TCACGAACGC GCTGGCTCTG
3451 GCGCTGCACG GCGCCCTTGT GCTCTCCGCG GCCAACCCCG AGAAGGGCAA
3501 GGAAATGCGG ACGCCGGATC ACGAGGATAC GTTCTTCCGC GATCTGGTCG
3551 GCTACTCGAT CGGGACGCTC GGCATCCACC GCCTCGGCCT GCTGCTCTCG
3601 CTGAGCGCCG TCTTCTTCAG CGCCCTCTGC ATGATCATTA CCGGCACCAT
3651 CTGGTTCGAT CAGTGGGTCG ACTGGTGGCA ATGGTGGGTG AAGCTGCCGT
3701 GGTGGGCGAA CATCCCGGGA GGCATCAATG GCTGAGTATC AGAACATCTT
3751 CTCCCAGGTC CAGGTCCGCG GACCGGCCGA CCTGGGGATG ACCGAAGACG
3801 TCAACCTGGC CAACCGTTCG GGCGTCGGTC CCTTCTCGAC CCTGCTCGGC
3851 TGGTTCGGCA ACGCCCAGCT CGGCCCGATC TATCTCGGCT CGCTCGGCGT
3901 CCTGTCCCTC TTCTCGGGCC TGATGTGGTT CTTCACCATC GGGATCTGGT
3951 TCTGGTATCA GGCGGGCTGG AACCCGGCCG TCTTCCTGCG CGACCTGTTC
4001 TTCTTCTCGC TCGAGCCGCC GGCACCCGAA TACGGTCTGT CCTTCGCGGC
4051 TCCGCTGAAG GAAGGCGGGC TGTGGCTGAT CGCGTCGTTC TTCATGTTCG
4101 TCGCGGTCTG GTCCTGGTGG GGCCGCACCT ATCTCCGCGC TCAGGCGCTG
4151 GGCATGGGCA AGCACACCGC CTGGGCGTTC CTCTCGGCCA TCTGGCTGTG
4201 GATGGTGCTG GGCTTCATCC GTCCGATCCT CATGGGGTCC TGGTCGGAAG
4251 CGGTTCCCTA CGGCATCTTC TCGCACCTCG ACTGGACGAA CAACTTCTCG
4301 CTCGTCCACG GCAACCTGTT CTACAACCCC TTCCACGGTC TCTCGATCGC
4351 CTTCCTCTAC GGGTCGGCCC TGCTCTTCGC GATGCACGGT GCGACCATCC
4401 TCGCGGTCTC CCGCTTCGGC GGCGAGCGCG AGCTGGAGCA GATCGCCGAC
4451 CGCGGGACGG CAGCGGAGCG GGCCGCCCTC TTCTGGCGCT GGACCATGGG
4501 TTTCAACGCC ACGATGGAAG GCATCCACCG CTGGGCCATC TGGATGGCGG
4551 TCCTCGTGAC CCTCACCGGC GGCATCGGGA TCCTGCTCTC GGGCACGGTC
4601 GTGGACAACT GGTACGTCTG GGGCCAGAAC CACGGCATGG CGCCGCTGAA
4651 CTGAGGAGCG ATCACAATGG CTGACAAGAC CATCTTCAAC GATCACCTCA
4701 ACACCAATCC GAAGACCAAC CTTCGCCTCT GGGTCGCTTT CCAGATGATG
4751 AAGGGTGCGG GCTGGGCTGG CGGCGTGTTC TTCGGGACGC TCCTTCTCAT
4801 CGGGTTCTTC CGGGTGGTCG GGCGGATGCT TCCGATCCAG GAGAACCAGG
4851 CTCCGGCGCC GAACATCACC GGCGCTCTGG AGACCGGGAT CGAGCTGATC
4901 AAGCATCTCG TCTGAGACAA GTCTCGGGGC AGGGCGGCGC GAGGCCGCCC
4951 GCTCCTCCAA GTCCGGGCCA TATCGCCGGC CCGGGTCCGG GGCGACACCA
5001 CAGCCCGGTT CCCTTCCTGT TGGCGACAGG GACCTGGTGC CGTGTGGAAG
5051 ACCGCACGGC ACCCTTTTGA CATTCACGGG AGGCTCTGAT GACCAATCCC

5101 ACCCCGCGAC CCGAAACCCC GCTTTTGGAT CGCGTCTGCT GCCCGGCCGA
5151 CATGAAGGCG CTGAGTGACG CCGAACTGGA GCGGCTGGCC GACGAAGTGC
5201 GTTCCGAGGT CAGTGATAGG GGTAGTTTCT TATTTTAGGC AGTTTATATG
5251 AAATTAAGAC ATGCAGATGT CACAGTGGAT ATTGAACTGG TCTCGAAAGC
5301 TCAATATCCC CCAAAGCACA AGCACAAACT TCGACATCAT GCAGAAGCGT
5351 TTCCCGAACC GCGTCTTCGA CGTGGGCATC GCCGAGCAGC ATGCCGTGAC
5401 CTTCGCGGCC GGCCTCGCCG GGGCCGGGAT GAAGCCCTTC TGCGCGATCT
5451 ATTCCTCGTT CCTGCAACGG GGTTACGACC AGATCGCCCA TGACGTGGCG
5501 CTGCAGAACC TTCCCGTCCG CTTCGTGATC GACCGGGCGG GGCTCGTGGG
5551 GGCCGATGGC GCGACCCATG CGGGGGCCTT CGACGTTGGC TTCATCACTT
5601 CGCTGCCCAA CATGACCGTG ATGGCCGCGG CCGACGAGGC CGAGCTCATC
5651 CACATGATCG CCACCGCCGT GGCCTTCGAC GAGGGCCCCA TCGCCTTCCG
5701 CTTCCCGCGG GGCGAGGGGG TGGGCGTCGA GATGCCCGAG CGCGGGACGG
5751 TGCTGGAGCC CGGCCGGGGC CGCGTGGTGC GCGAAGGGAC GGATGTCGCG
5801 ATCCTCTCCT TCGGCGCGCA TCTGCACGAG GCCTTGCAGG CGGCGAAACT
5851 TCTCGAGGCC GAGGGGGTGA GCGTGACCGT GGCCGACGCC CGCTTCTCGC
5901 GCCCGCTCGA CACGGGGCTC ATCGACCAGC TCGTGCGCCA TCACGCGGCG
5951 CTGGTAACGG TGGAGCAGGG GGCCATGGGC GGCTTCGGCG CCCATGTCAT
6001 GCACTATCTC GCCAATTCCG GCGGCTTCGA CGGGGGCCTC GCGCTCCGGG
6051 TCATGACGCT GCCCGACCGC TTCATCGAGC AGGCGAGCCC CGAGGACATG
6101 TATGCCGATG CGGGGCTGCG GGCCGAGGAT ATCAAGCTTG GCGTAATCAT
6151 GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATTCCACAC
6201 AACATACGAG CCGGAAGCAT AAAGTGTAAA GCCTGGGGTG CCTAATGAGT
6251 GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG
6301 GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA
6351 GGCGGTTTGC GTATTGGGCG CTCGGTCTTG CCTTGCTCGT CGGTGATGTA
6401 CTTCACCAGC TCCGCGAAGT CGCTCTTCTT GATGGAGCGC ATGGGGACGT
6451 GCTTGGCAAT CACGCGCACC CCCCGGCCGT TTTAGCGGCT AAAAAAGTCA
6501 TGGCTCTGCC CTCGGGCGGA CCACGCCCAT CATGACCTTG CCAAGCTCGT
6551 CCTGCTTCTC TTCGATCTTC GCCAGCAGGG CGAGGATCGT GGCATCACCG
6601 AACCGCGCCG TGCGCGGGTC GTCGGTGAGC CAGAGTTTCA GCAGGCCGCC
6651 CAGGCGGCCC AGGTCGCCAT TGATGCGGGC CAGCTCGCGG ACGTGCTCAT
6701 AGTCCACGAC GCCCGTGATT TTGTAGCCCT GGCCGACGGC CAGCAGGTAG
6751 GCCGACAGGC TCATGCCGGC CGCCGCCGCC TTTTCCTCAA TCGCTCTTCG
```

-continued

```
 6801 TTCGTCTGGA AGGCAGTACA CCTTGATAGG TGGGCTGCCC TTCCTGGTTG
 6851 GCTTGGTTTC ATCAGCCATC CGCTTGCCCT CATCTGTTAC GCCGGCGGTA
 6901 GCCGGCCAGC CTCGCAGAGC AGGATTCCCG TTGAGCACCG CCAGGTGCGA
 6951 ATAAGGGACA GTGAAGAAGG AACACCCGCT CGCGGGTGGG CCTACTTCAC
 7001 CTATCCTGCC CGGCTGACGC CGTTGGATAC ACCAAGGAAA GTCTACACGA
 7051 ACCCTTTGGC AAAATCCTGT ATATCGTGCG AAAAAGGATG GATATACCGA
 7101 AAAAATCGCT ATAATGACCC CGAAGCAGGG TTATGCAGCG GAAAAGCGCC
 7151 ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT
 7201 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGGAAAC GCCTGGTATC
 7251 TTTATAGTCC TGTCGGGTTT CGCCACCTCT GACTTGAGCG TCGATTTTTG
 7301 TGATGCTCGT CAGGGGGGCG GAGCCTATGG AAAAACGCCA GCAACGCGGC
 7351 CTTTTTACGG TTCCTGGCCT TTTGCTGGCC TTTTGCTCAC ATGTTCTTTC
 7401 CTGCGTTATC CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA
 7451 GCTGATACCG CTCGCCGCAG CCGAACGACC GAGCGCAGCG AGTCAGTGAG
 7501 CGAGGAAGCG GAAGAGCGCC AGAAGGCCGC CAGAGAGGCC GAGCGCGGCC
 7551 GTGAGGCTTG GACGCTAGGG CAGGGCATGA AAAAGCCCGT AGCGGGCTGC
 7601 TACGGGCGTC TGACGCGGTG GAAAGGGGGA GGGGATGTTG TCTACATGGC
 7651 TCTGCTGTAG TGAGTGGGTT GCGCTCCGGC AGCGGTCCTG ATCAATCGTC
 7701 ACCCTTTCTC GGTCCTTCAA CGTTCCTGAC AACGAGCCTC CTTTTCGCCA
 7751 ATCCATCGAC AATCACCGCG AGTCCCTGCT CGAACGCTGC GTCCGGACCG
 7801 GCTTCGTCGA AGGCGTCTAT CGCGGCCCGC AACAGCGGCG AGAGCGGAGC
 7851 CTGTTCAACG GTGCCGCCGC GCTCGCCGGC ATCGCTGTCG CCGGCCTGCT
 7901 CCTCAAGCAC GGCCCCAACA GTGAAGTAGC TGATTGTCAT CAGCGCATTG
 7951 ACGGCGTCCC CGGCCGAAAA ACCCGCCTCG CAGAGGAAGC GAAGCTGCGC
 8001 GTCGGCCGTT TCCATCTGCG GTGCGCCCGG TCGCGTGCCG GCATGGATGC
 8051 GCGCGCCATC GCGGTAGGCG AGCAGCGCCT GCCTGAAGCT GCGGGCATTC
 8101 CCGATCAGAA ATGAGCGCCA GTCGTCGTCG GCTCTCGGCA CCGAATGCGT
 8151 ATGATTCTCC GCCAGCATGG CTTCGGCCAG TGCGTCGAGC AGCGCCCGCT
 8201 TGTTCCTGAA GTGCCAGTAA AGCGCCGGCT GCTGAACCCC CAACCGTTCC
 8251 GCCAGTTTGC GTGTCGTCAG ACCGTCTACG CCGACCTCGT TCAACAGGTC
 8301 CAGGGCGGCA CGGATCACTG TATTCGGCTG CAACTTTGTC ATGCTTGACA
 8351 CTTTATCACT GATAAACATA ATATGTCCAC CAACTTATCA GTGATAAAGA
 8401 ATCCGCGCGT TCAATCGGAC CAGCGGAGGC TGGTCCGGAG GCCAGACATG
 8451 AAACCCAACA TACCCCTGAT CGTAATTCTG AGCACTGTCG CGCTCGACGC

 8501 TGTCGGCATC GGCCTGATTA TGCCGGTGCT GCCGGGCCTC CTGCGCGATC
 8551 TGGTTCACTC GAACGACGTC ACCGCCCACT ATGGCATTCT GCTGGCGCTG
 8601 TATGCGTTGG TGCAATTTGC CTGCGCACCT GTGCTGGGCG CGCTGTCGGA
 8651 TCGTTTCGGG CGGCGGCCAA TCTTGCTCGT CTCGCTGGCC GGCGCCACTG
 8701 TCGACTACGC CATCATGGCG ACAGCGCCTT TCCTTTGGGT TCTCTATATC
 8751 GGGCGGATCG TGGCCGGCAT CACCGGGGCG ACTGGGGCGG TAGCCGGCGC
 8801 TTATATTGCC GATATCACTG ATGGCGATGA GCGCGCGCGG CACTTCGGCT
 8851 TCATGAGCGC CTGTTTCGGG TTCGGGATGG TCGCGGGACC TGTGCTCGGT
 8901 GGGCTGATGG GCGGTTTCTC CCCCCACGCT CCGTTCTTCG CCGCGGCAGC
 8951 CTTGAACGGC CTCAATTTCC TGACGGGCTG TTTCCTTTTG CCGGAGTCGC
 9001 ACAAAGGCGA ACGCCGGCCG TTACGCCGGG AGGCTCTCAA CCCGCTCGCT
 9051 TCGTTCCGGT GGGCCCGGGG CATGACCGTC GTCGCCGCCC TGATGGCGGT
 9101 CTTCTTCATC ATGCAACTTG TCGGACAGGT GCCGGCCGCG CTTTGGGTCA
 9151 TTTTCGGCGA GGATCGCTTT CACTGGGACG CGACCACGAT CGGCATTTCG
 9201 CTTGCCGCAT TTGGCATTCT GCATTCACTC GCCCAGGCAA TGATCACCGG
 9251 CCCTGTAGCC GCCCGGCTCG GCGAAAGGCG GGCACTCATG CTCGGAATGA
 9301 TTGCCGACGG CACAGGCTAC ATCCTGCTTG CCTTCGCGAC ACGGGGATGG
 9351 ATGGCGTTCC CGATCATGGT CCTGCTTGCT TCGGGTGGCA TCGGAATGCC
 9401 GGCGCTGCAA GCAATGTTGT CCAGGCAGGT GGATGAGGAA CGTCAGGGGC
 9451 AGCTGCAAGG CTCACTGGCG GCGCTCACCA GCCTGACCTC GATCGTCGGA
 9501 CCCCTCCTCT TCACGGCGAT CTATGCGGCT TCTATAACAA CGTGGAACGG
 9551 GTGGGCATGG ATTGCAGGCG CTGCCCTCTA CTTGCTCTGC CTGCCGGCGC
 9601 TGCGTCGCGG GCTTTGGAGC GGCGCAGGGC AACGAGCCGA TCGCTGATCG
 9651 TGGAAACGAT AGGCCTATGC CATGCGGGTC AAGGCGACTT CCGGCAAGCT
 9701 ATACGCGCCC TAGGAGTGCG GTTGGAACGT TGGCCCAGCC AGATACTCCC
 9751 GATCACGAGC AGGACGCCGA TGATTTGAAG CGCACTCAGC GTCGATCCA
 9801 AGAACAACCA TCCTAGCAAC ACGGCGGTCC CCGGGCTGAG AAAGCCCAGT
 9851 AAGGAAACAA CTGTAGGTTC GAGTCGCGAG ATCCCCCGGA ACCAAAGGAA
 9901 GTAGGTTAAA CCCGCTCCGA TCAGGCCGAG CCACGCCAGG CCGAGAACAT
 9951 TGGTTCCTGT AGGCATCGGG ATTGGCGGAT CAAACACTAA AGCTACTGGA
10001 ACGAGCAGAA GTCCTCCGGC CGCCAGTTGC CAGGCGGTAA AGGTGAGCAG
10051 AGGCACGGGA GGTTGCCACT TGCGGGTCAG CACGGTTCCG AACGCCATGG
10101 AAACCGCCCC CGCCAGGCCC GCTGCGACGC CGACAGGATC TAGCGCTGCG
10151 TTTGGTGTCA ACACCAACAG CGCCACGCCC GCAGTTCCGC AAATAGCCCC
```

-continued

```
10201 CAGGACCGCC ATCAATCGTA TCGGGCTACC TAGCAGAGCG GCAGAGATGA
10251 ACACGACCAT CAGCGGCTGC ACAGCGCCTA CCGTCGCCGC GACCCCGCCC
10301 GGCAGGCGGT AGACCGAAAT AAACAACAAG CTCCAGAATA GCGAAATATT
10351 AAGTGCGCCG AGGATGAAGA TGCGCATCCA CCAGATTCCC GTTGGAATCT
10401 GTCGGACGAT CATCACGAGC AATAAACCCG CCGGCAACGC CCGCAGCAGC
10451 ATACCGGCGA CCCCTCGGCC TCGCTGTTCG GGCTCCACGA AAACGCCGGA
10501 CAGATGCGCC TTGTGAGCGT CCTTGGGGCC GTCCTCCTGT TTGAAGACCG
10551 ACAGCCCAAT GATCTCGCCG TCGATGTAGG CGCCGAATGC CACGGCATCT
10601 CGCAACCGTT CAGCGAACGC CTCCATGGGC TTTTTCTCCT CGTGCTCGTA
10651 AACGGACCCG AACATCTCTG GAGCTTTCTT CAGGGCCGAC AATCGGATCT
10701 CGCGGAAATC CTGCACGTCG GCCGCTCCAA GCCGTCGAAT CTGAGCCTTA
10751 ATCACAATTG TCAATTTTAA TCCTCTGTTT ATCGGCAGTT CGTAGAGCGC
10801 GCCGTGCGTC CCGAGCGATA CTGAGCGAAG CAAGTGCGTC GAGCAGTGCC
10851 CGCTTGTTCC TGAAATGCCA GTAAAGCGCT GGCTGCTGAA CCCCCAGCCG
10901 GAACTGACCC CACAAGGCCC TAGCGTTTGC AATGCACCAG GTCATCATTG
10951 ACCCAGGCGT GTTCCACCAG GCCGCTGCCT CGCAACTCTT CGCAGGCTTC
11001 GCCGACCTGC TCGCGCCACT TCTTCACGCG GGTGGAATCC GATCCGCACA
11051 TGAGGCGGAA GGTTTCCAGC TTGAGCGGGT ACGGCTCCCG GTGCGAGCTG
11101 AAATAGTCGA ACATCCGTCG GGCCGTCGGC GACAGCTTGC GGTACTTCTC
11151 CCATATGAAT TTCGTGTAGT GGTCGCCAGC AAACAGCACG ACGATTTCCT
11201 CGTCGATCAG GACCTGGCAA CGGGACGTTT TCTTGCCACG GTCCAGGACG
11251 CGGAAGCGGT GCAGCAGCGA CACCGATTCC AGGTGCCCAA CGCGGTCGGA
11301 CGTGAAGCCC ATCGCCGTCG CCTGTAGGCG CGACAGGCAT TCCTCGGCCT
11351 TCGTGTAATA CCGGCCATTG ATCGACCAGC CCAGGTCCTG GCAAAGCTCG
11401 TAGAACGTGA AGGTGATCGG CTCGCCGATA GGGGTGCGCT TCGCGTACTC
11451 CAACACCTGC TGCCACACCA GTTCGTCATC GTCGGCCCGC AGCTCGACGC
11501 CGGTGTAGGT GATCTTCACG TCCTTGTTGA CGTGGAAAAT GACCTTGTTT
11551 TGCAGCGCCT CGCGCGGGAT TTTCTTGTTG CGCGTGGTGA ACAGGGCAGA
11601 GCGGGCCGTG TCGTTTGGCA TCGCTCGCAT CGTGTCCGGC CACGGCGCAA
11651 TATCGAACAA GGAAAGCTGC ATTTCCTTGA TCTGCTGCTT CGTGTGTTTC
11701 AGCAACGCGG CCTGCTTGGC CTCGCTGACC TGTTTTGCCA GGTCCTCGCC
11751 GGCGGTTTTT CGCTTCTTGG TCGTCATAGT TCCTCGCGTG TCGATGGTCA
11801 TCGACTTCGC CAAACCTGCC GCCTCCTGTT CGAGACGACG CGAACGCTCC
11851 ACGGCGGCCG ATGGCGCGGG CAGGGCAGGG GGAGCCAGTT GCACGCTGTC

11901 GCGCTCGATC TTGGCCGTAG CTTGCTGGAC CATCGAGCCG ACGGACTGGA
11951 AGGTTTCGCG GGGCGCACGC ATGACGGTGC GGCTTGCGAT GGTTTCGGCA
12001 TCCTCGGCGG AAAACCCCGC GTCGATCAGT TCTTGCCTGT ATGCCTTCCG
12051 GTCAAACGTC CGATTCATTC ACCCTCCTTG CGGGATTGCC CCGACTCACG
12101 CCGGGGCAAT GTGCCCTTAT TCCTGATTTG ACCCGCCTGG TGCCTTGGTG
12151 TCCAGATAAT CCACCTTATC GGCAATGAAG TCGGTCCCGT AGACCGTCTG
12201 GCCGTCCTTC TCGTACTTGG TATTCCGAAT CTTGCCCTGC ACGAATACCA
12251 GCGACCCCTT GCCCAAATAC TTGCCGTGGG CCTCGGCCTG AGAGCCAAAA
12301 CACTTGATGC GGAAGAAGTC GGTGCGCTCC TGCTTGTCGC CGGTCGTGGC
12351 CGCGCCAACC TTTGCGATCC GCAAGCGCGC GGTCGCCATC TTCACGCTGG
12401 AACAGTACGT CGAGGCGGGC ATCATGACCC GCGAGCAATA CGAGGTCATT
12451 AAAAGCGCCG TGATTGATGA TATAGCGGCC CGGCTGCTCC TGGTTCTCGC
12501 GCACCGAAAT GGGTGACTTC ACCCCGCGCT CTTTGATCGT GGCACCGATT
12551 TCCGCGATGC TCTCCGGGGA AAAGCCGGGG TTGTCGGCCG TCCGCGGCTG
12601 ATGCGGATCT TCGTCGATCA GGTCCAGGTC CAGCTCGATA GGGCCGGAAC
12651 CGCCCTGAGA CGCCGCAGGA GCGTCCAGGA GGCTCGACAG GTCGCCGATG
12701 CTATCCAACC CCAGGCCGGA CGGCTGCGCC GCGCCTGCGG CTTCCTGAGC
12751 GGCCGCAGCG GTGTTTTTCT TGGTGGTCTT GGCTTGAGCC GCAGTCATTG
12801 GGAAATCTCC ATCTTCGTGA ACACGTAATC AGCCAGGGCG CGAACCTCTT
12851 TCGATGCCTT GCGCGCGGCC GTTTTCTTGA TCTTCCAGAC CGGCACACCG
12901 GATGCGAGGG CATCGGCGAT GCTGCTGCGC AGGCCAACGG TGGCCGGAAT
12951 CATCATCTTG GGGTACGCGG CCAGCAGCTC GGCTTGGTGG CGCGCGTGGC
13001 GCGGATTCCG CGCATCGACC TTGCTGGGCA CCATGCCAAG GAATTGCAGC
13051 TTGGCGTTCT TCTGGCGCAC GTTCGCAATG GTCGTGACCA TCTTCTTGAT
13101 GCCCTGGATG CTGTACGCCT CAAGCTCGAT GGGGGACAGC ACATAGTCGG
13151 CCGCGAAGAG GGCGGCCGCC AGGCCGACGC CAAGGGTCGG GGCCGTGTCG
13201 ATCAGGCACA CGTCGAAGCC TTGGTTCGCC AGGGCCTTGA TGTTCGCCCC
13251 GAACAGCTCG CGGGCGTCGT CCAGCGACAG CCGTTCGGCG TTCGCCAGTA
13301 CCGGGTTGGA CTCGATGAGG GCGAGGCGCG CGGCCTGGCC GTCGCCGGCT
13351 GCGGGTGCGG TTTCGGTCCA GCCGCCGGCA GGGACAGCGC CGAACAGCTT
13401 GCTTGCATGC AGGCCGGTAG CAAAGTCCTT GAGCGTGTAG GACGCATTGC
13451 CCTGGGGGTC CAGGTCGATC ACGGCAACCC GCAAGCCGCG CTCGAAAAAG
13501 TCGAAGGCAA GATGCACAAG GGTCGAAGTC TTGCCGACGC CGCCTTTCTG
13551 GTTGGCCGTG ACCAAAGTTT TCATCGTTTG GTTTCCTGTT TTTTCTTGGC
```

-continued

```
13601 GTCCGCTTCC CACTTCCGGA CGATGTACGC CTGATGTTCC GGCAGAACCG
13651 CCGTTACCCG CGCGTACCCC TCGGGCAAGT TCTTGTCCTC GAACGCGGCC
13701 CACACGCGAT GCACCGCTTG CGACACTGCG CCCCTGGTCA GTCCCAGCGA
13751 CGTTGCGAAC GTCGCCTGTG GCTTCCCATC GACTAAGACG CCCCGCGCTA
13801 TCTCGATGGT CTGCTGCCCC ACTTCCAGCC CCTGGATCGC CTCCTGGAAC
13851 TGGCTTTCGG TAAGCCGTTT CTTCATGGAT AACACCCATA ATTTGCTCCG
13901 CGCCTTGGTT GAACATAGCG GTGACAGCCG CCAGCACATG AGAGAAGTTT
13951 AGCTAAACAT TTCTCGCACG TCAACACCTT TAGCCGCTAA AACTCGTCCT
14001 TGGCGTAACA AAACAAAAGC CCGGAAACCG GGCTTTCGTC TCTTGCCGCT
14051 TATGGCTCTG CACCCGGCTC CATCACCAAC AGGTCGCGCA CGCGCTTCAC
14101 TCGGTTGCGG ATCGACACTG CCAGCCCAAC AAAGCCGGTT GCCGCCGCCG
14151 CCAGGATCGC GCCGATGATG CCGGCCACAC CGGCCATCGC CCACCAGGTC
14201 GCCGCCTTCC GGTTCCATTC CTGCTGGTAC TGCTTCGCAA TGCTGGACCT
14251 CGGCTCACCA TAGGCTGACC GCTCGATGGC GTATGCCGCT TCTCCCCTTG
14301 GCGTAAAACC CAGCGCCGCA GGCGGCATTG CCATGCTGCC CGCCGCTTTC
14351 CCGACCACGA CGCGCGCACC AGGCTTGCGG TCCAGACCTT CGGCCACGGC
14401 GAGCTGCGCA AGGACATAAT CAGCCGCCGA CTTGGCTCCA CGCGCCTCGA
14451 TCAGCTCTTG CACTCGCGCG AAATCCTTGG CCTCCACGGC CGCCATGAAT
14501 CGCGCACGCG GCGAAGGCTC CGCAGGGCCG
```

PRKSSHT1DPUF.SEQ

```
    1 CCACCCAGGC CGCCGCCCTC ACTGCCCGGC ACCTGGTCGC TGAATGTCGA
   51 TGCCAGCACC TGCGGCACGT CAATGCTTCC GGGCGTCGCG CTCGGGCTGA
  101 TCGCCCATCC CGTTACTGCC CCGATCCCGG CAATGGCAAG GACTGCCAGC
  151 GCTGCCATTT TTGGGGTGAG GCCGTTCGCG GCCGAGGGGC GCAGCCCCTG
  201 GGGGGATGGG AGGCCCGCGT TAGCGGGCCG GGAGGGTTCG AGAAGGGGGG
  251 GCACCCCCCT TCGGCGTGCG CGGTCACGCG CACAGGGCGC AGCCCTGGTT
  301 AAAAACAAGG TTTATAAATA TTGGTTTAAA AGCAGGTTAA AAGACAGGTT
  351 AGCGGTGGCC GAAAAACGGG CGGAAACCCT TGCAAATGCT GGATTTTCTG
  401 CCTGTGGACA GCCCCTCAAA TGTCAATAGG TGCGCCCCTC ATCTGTCAGC
  451 ACTCTGCCCC TCAAGTGTCA AGGATCGCGC CCCTCATGTG TCAGTAGTCG
  501 CGCCCCTCAA GTGTCAATAC CGCAGGGCAC TTATCCCCAG GCTTGTCCAC
  551 ATCATCTGTG GGAAACTCGC GTAAAATCAG GCGTTTTCGC CGATTTGCGA
  601 GGCTGGCCAG CTCCACGTCG CCGGCCGAAA TCGAGCCTGC CCCTCATCTG
  651 TCAACGCCGC GCCGGGTGAG TCGGCCCCTC AAGTGTCAAC GTCCGCCCCT
  701 CATCTGTCAG TGAGGGCCAA GTTTTCCGCG AGGTATCCAC AACGCCGGCG
  751 GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC GCGTTTGCAG
  801 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG CCCGGTGAGC
  851 GTCGGAAAGG CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG
  901 TCGTTCGGCT GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG
  951 TTATCCACAG AATCAGGGGA TAACGCAGGA AAGAACATGT GAGCAAAAGG
 1001 CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG GCGTTTTTCC
 1051 ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
 1101 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
 1151 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC
 1201 TGTCCGCCTT TCTCCCTTCG GGAAGCGTGG CGCCATTCGC CATTCAGGCT
 1251 GCGCAACTGT TGGGAAGGGC GATCGGTGCG GGCCTCTTCG CTATTACGCC
 1301 AGCTGGCGAA AGGGGGATGT GCTGCAAGGC GATTAAGTTG GGTAACGCCA
 1351 GGGTTTTCCC AGTCACGACG TTGTAAAACG ACGGCCAGTG AATTCGGCCG
 1401 CGGGCTGGCC GAGGTGCTGG GCAAGCCCTA CCTCCAGGCC CCCATCGGGG
 1451 TCGAGAGCAC GACCGCCTTC CTGCGCCGCC TGGGCGAGAT TCTGGGCCTC
 1501 GATCCGGAGC CCTTCATCGA GCGCGAGAAG CACTCGACGC TGAAGCCCGT
 1551 GTGGGATCTG TGGCGGAGTG TCACGCAGGA CTTCTTCGGG ACGGCCAATT
 1601 TCGGAATCGT GGCGACCGAA ACTTATGCAA GAGGCATCCG AAACTATCTC
 1651 GAAGGCGATC TCGGGCTGCC CTGCGCCTTC GCCGTGGCCC GCAAGAGGGG
```

-continued

```
1701 CTCGAAGACC GACAACGAAG CGGTGCGCGG ACTGATCCGC CAGCACCGTC
1751 CGCTCGTGCT CATGGGGTCG ATCAACGAGA AGATTTACCT TGCGGAACTG
1801 AAAGCCGGTC ACGGCCCGCA ACCCTCTTTC ATCGCTGCCT CTTTCCCGGG
1851 TGCGGCGATC CGGCGCGCTA CCGGAACGCC CGTTATGGGA TATGCAGGTG
1901 CTACGTGGTT ACTGCAGGAA GTTTGCAACG CCCTGTTCGA CGCCCTGTTC
1951 CACATTCTGC CCCTCGGGAC GGAGATGGAC AGCGCCGCCG CCACACCGAC
2001 GACACTGCGC CGCGACTTCC CGTGGGATGC CGATGCGCAA gcGGCCCTGG
2051 ACCGCATCGT AGAGGAGCAT CCGGTTCTCA CCCGGATCAG CGCCGCGCGT
2101 GCCTTGCGCG ACGCCGCCGA GAAGGCTGCC CTCGATGCCG GTGCCGAGAG
2151 GGTCGTGAGA GAGACTGTCG AAGCCCTGCG TGGGCCGGGC TTCGGCGAGA
2201 GGAAGGGAGA GAACCAATGA GCGATCATGC CGTCAACACG CCGGTCCATG
2251 CCGCCAGGGC CCACGGGCAC CGAGCACCAC GTGCCGAGTT CTACGTCTAC
2301 TTCGCCGTCA TTCTGCTGGG CGCCTTCCCG GTGGCCTTCG TGAGCTGGAT
2351 CGTCTCGACG ATCCGCCACC GCAGGCTTCC CAAGCGCGGC CCCTTCGCGT
2401 CCGCCTGGTT CGATGCCAAG GCGATCACGC CGCTGATTTT CCGCGCCTGA
2451 CCGCAGGTCA GGTTGCGACA CGCCATTCGT CGTCTCCCCA AGGGGCGGCG
2501 GATTAATCGG GAGGGCATGG TGCCTTACCG TAACCCACGC CACCAGCATG
2551 TGGCGACTAG GGGAGGATAG TAATGAAGTT CCAAGTCAAG GCCCTCGCCG
2601 CCATCGCCGC ATTCGCGGCG CTGCCGGCGC TCGCGCAGAC TAGTTCCATA
2651 TGATAGATCT CACGACTAGC ACGACTACCA GTAATAGGCC GGCCCTCCGT
2701 CGCGGGCGGC ACCCACGCCC GCATCGATTC CAAGGTTCAG CCATTGAGAC
2751 GGCTCCGCTT CGCGCGCAAG CGCGGGTTGG GCCGACTGCA AGCGGAGAGG
2801 GAAGCATGGC ACTGCTCAGC TTCGAGCGAA AATATCGCGT GCCGGGGGGC
2851 ACGCTGGTCG GCGGAAACCT GTTCGACTTC TGGGTCGGCC CTTTCTATGT
2901 CGGCTTCTTC GGGGTTGCGA CGTTTTTCTT CGCGGCCCTG GGTATCATTC
2951 TGATTGCCTG GAGTGCCGTA CTCCAGGGTA CCTGGAACCC CCAACTCATC
3001 TCTGTCTACC CGCCGGCCCT TGAATATGGC CTGGGAGGTG CACCCCTCGC
3051 AAAAGGCGGG CTGTGGCAGA TCATCACGAT CTGCGCCACT GGTGCCTTCG
3101 TCAGCTGGGC GCTGCGCGAA GTCGAAATCT GCCGTAAGCT GGGCATCGGG
3151 TACCACATCC CGTTCGCCTT CGCGTTCGCC ATCCTGGCCT ACCTGACGCT
3201 GGTGCTGTTC CGCCCGGTGA TGATGGGCGC CTGGGGCTAT GCCTTCCCCT
3251 ACGGGATCTG GACGCACCTC GACTGGGTGT CGAACACGGG CTACACCTAC
3301 GGCAACTTCC ACTACAACCC TGCCCACATG ATCGCCATCT CGTTCTTCTT
3351 CACGAACGCG CTGGCTCTGG CGCTGCACGG CGCCCTTGTG CTCTCCGCGG

3401 CCAACCCCGA GAAGGGCAAG GAAATGCGGA CGCCGGATCA CGAGGATACG
3451 TTCTTCCGCG ATCTGGTCGG CTACTCGATC GGGACGCTCG GCATCCACCG
3501 CCTCGGCCTG CTGCTCTCGC TGAGCGCCGT CTTCTTCAGC GCCCTCTGCA
3551 TGATCATTAC CGGCACCATC TGGTTCGATC AGTGGGTCGA CTGGTGGCAA
3601 TGGTGGGTGA AGCTGCCGTG GTGGGCGAAC ATCCCGGGAG GCATCAATGG
3651 CTGAGTATCA GAACATCTTC TCCCAGGTCC AGGTCCGCGG ACCGGCCGAC
3701 CTGGGGATGA CCGAAGACGT CAACCTGGCC AACCGTTCGG GCGTCGGTCC
3751 CTTCTCGACC CTGCTCGGCT GGTTCGGCAA CGCCCAGCTC GGCCCGATCT
3801 ATCTCGGCTC GCTCGGCGTC CTGTCCCTCT TCTCGGGCCT GATGTGGTTC
3851 TTCACCATCG GGATCTGGTT CTGGTATCAG GCGGGCTGGA ACCCGGCCGT
3901 CTTCCTGCGC GACCTGTTCT TCTTCTCGCT CGAGCCGCCG GCACCCGAAT
3951 ACGGTCTGTC CTTCGCGGCT CCGCTGAAGG AAGGCGGGCT GTGGCTGATC
4001 GCGTCGTTCT TCATGTTCGT CGCGGTCTGG TCCTGGTGGG GCCGCACCTA
4051 TCTCCGCGCT CAGGCGCTGG GCATGGGCAA GCACACCGCC TGGGCGTTCC
4101 TCTCGGCCAT CTGGCTGTGG ATGGTGCTGG GCTTCATCCG TCCGATCCTC
4151 ATGGGGTCCT GGTCGGAAGC GGTTCCCTAC GGCATCTTCT CGCACCTCGA
4201 CTGGACGAAC AACTTCTCGC TCGTCCACGG CAACCTGTTC TACAACCCCT
4251 TCCACGGTCT CTCGATCGCC TTCCTCTACG GGTCGGCCCT GCTCTTCGCG
4301 ATGCACGGTG CGACCATCCT CGCGGTCTCC CGCTTCGGCG GCGAGCGCGA
4351 GCTGGAGCAG ATCGCCGACC GCGGGACGGC AGCGGAGCGG GCCGCCCTCT
4401 TCTGGCGCTG GACCATGGGT TTCAACGCCA CGATGGAAGG CATCCACCGC
4451 TGGGCCATCT GGATGGCGGT CCTCGTGACC CTCACCGGCG GCATCGGGAT
4501 CCTGCTCTCG GGCACGGTCG TGGACAACTG GTACGTCTGG GGCCAGAACC
4551 ACGGCATGGC GCCGCTGAAC TGAGGAGCGA TCACAATGGC TGACAAGACC
4601 ATCTTCAACG ATCACCTCAA CACCAATCCG AAGACCAACC TTCGCCTCTG
4651 GGTCGCTTTC CAGATGATGA AGGGTGCGGG CTGGGCTGGC GGCGTGTTCT
4701 TCGGGACGCT CCTTCTCATC GGGTTCTTCC GGGTGGTCGG GCGGATGCTT
4751 CCGATCCAGG AGAACCAGGC TCCGGCGCCG AACATCACCG GCGCTCTGGA
4801 GACCGGGATC GAGCTGATCA AGCATCTCGT CTGAGACAAG TCTCGGGGCA
4851 GGGCGGCGCG AGGCCGCCCG CTCCTCCAAG TCCGGGCCAT ATCGCCGGCC
4901 CGGGTCCGGG GCGACACCAC AGCCCGGTTC CCTTCCTGTT GGCGACAGGG
4951 ACCTGGTGCC GTGTGGAAGA CCGCACGGCA CCCTTTTGAC ATTCACGGGA
5001 GGCTCTGATG ACCAATCCCA CCCCGCGACC CGAAACCCCG CTTTTGGATC
5051 GCGTCTGCTG CCCGGCCGAC ATGAAGGCGC TGAGTGACGC CGAACTGGAG
```

-continued

```
5101 CGGCTGGCCG ACGAAGTGCG TTCCGAGGTC AGTGATAGGG GTAGTTTCTT
5151 ATTTTAGGCA GTTTATATGA AATTAAGACA TGCAGATGTC ACAGTGGATA
5201 TTGAACTGGT CTCGAAAGCT CAATATCCCC CAAAGCACAA GCACAAACTT
5251 CGACATCATG CAGAAGCGTT TCCCGAACCG CGTCTTCGAC GTGGGCATCG
5301 CCGAGCAGCA TGCCGTGACC TTCGCGGCCG GCCTCGCCGG GGCCGGGATG
5351 AAGCCCTTCT GCGCGATCTA TTCCTCGTTC CTGCAACGGG GTTACGACCA
5401 GATCGCCCAT GACGTGGCGC TGCAGAACCT TCCCGTCCGC TTCGTGATCG
5451 ACCGGGCGGG GCTCGTGGGG GCCGATGGCG CGACCCATGC GGGGGCCTTC
5501 GACGTTGGCT TCATCACTTC GCTGCCCAAC ATGACCGTGA TGGCCGCGGC
5551 CGACGAGGCC GAGCTCATCC ACATGATCGC CACCGCCGTG GCCTTCGACG
5601 AGGGCCCCAT CGCCTTCCGC TTCCCGCGGG GCGAGGGGGT GGGCGTCGAG
5651 ATGCCCGAGC GCGGGACGGT GCTGGAGCCC GGCCGGGGCC GCGTGGTGCG
5701 CGAAGGGACG GATGTCGCGA TCCTCTCCTT CGGCGCGCAT CTGCACGAGG
5751 CCTTGCAGGC GGCGAAACTT CTCGAGGCCG AGGGGGTGAG CGTGACCGTG
5801 GCCGACGCCC GCTTCTCGCG CCCGCTCGAC ACGGGGCTCA TCGACCAGCT
5851 CGTGCGCCAT CACGCGGCGC TGGTAACGGT GGAGCAGGGG GCCATGGGCG
5901 GCTTCGGCGC CCATGTCATG CACTATCTCG CCAATTCCGG CGGCTTCGAC
5951 GGGGGCCTCG CGCTCCGGGT CATGACGCTG CCCGACCGCT TCATCGAGCA
6001 GGCGAGCCCC GAGGACATGT ATGCCGATGC GGGGCTGCGG GCCGAGGATA
6051 TCAAGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA
6101 TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG
6151 CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAATTGC GTTGCGCTCA
6201 CTGCCCGCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT
6251 CGGCCAACGC GCGGGGAGAG GCGGTTTGCG TATTGGGCGC TCGGTCTTGC
6301 CTTGCTCGTC GGTGATGTAC TTCACCAGCT CCGCGAAGTC GCTCTTCTTG
6351 ATGGAGCGCA TGGGGACGTG CTTGGCAATC ACGCGCACCC CCCGGCCGTT
6401 TTAGCGGCTA AAAAAGTCAT GGCTCTGCCC TCGGGCGGAC CACGCCCATC
6451 ATGACCTTGC CAAGCTCGTC CTGCTTCTCT TCGATCTTCG CCAGCAGGGC
6501 GAGGATCGTG GCATCACCGA ACCGCGCCGT GCGCGGGTCG TCGGTGAGCC
6551 AGAGTTTCAG CAGGCCGCCC AGGCGGCCCA GGTCGCCATT GATGCGGGCC
6601 AGCTCGCGGA CGTGCTCATA GTCCACGACG CCCGTGATTT TGTAGCCCTG
6651 GCCGACGGCC AGCAGGTAGG CCGACAGGCT CATGCCGGCC GCCGCCGCCT
6701 TTTCCTCAAT CGCTCTTCGT TCGTCTGGAA GGCAGTACAC CTTGATAGGT
6751 GGGCTGCCCT TCCTGGTTGG CTTGGTTTCA TCAGCCATCC GCTTGCCCTC

6801 ATCTGTTACG CCGGCGGTAG CCGGCCAGCC TCGCAGAGCA GGATTCCCGT
6851 TGAGCACCGC CAGGTGCGAA TAAGGGACAG TGAAGAAGGA ACACCCGCTC
6901 GCGGGTGGGC CTACTTCACC TATCCTGCCC GGCTGACGCC GTTGGATACA
6951 CCAAGGAAAG TCTACACGAA CCCTTTGGCA AAATCCTGTA TATCGTGCGA
7001 AAAAGGATGG ATATACCGAA AAAATCGCTA TAATGACCCC GAAGCAGGGT
7051 TATGCAGCGG AAAAGCGCCA CGCTTCCCGA AGGGAGAAAG GCGGACAGGT
7101 ATCCGGTAAG CGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGCTTCCA
7151 GGGGGAAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC GCCACCTCTG
7201 ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA
7251 AAAACGCCAG CAACGCGGCC TTTTTACGGT TCCTGGCCTT TTGCTGGCCT
7301 TTTGCTCACA TGTTCTTTCC TGCGTTATCC CCTGATTCTG TGGATAACCG
7351 TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCAGC CGAACGACCG
7401 AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCCA GAAGGCCGCC
7451 AGAGAGGCCG AGCGCGGCCG TGAGGCTTGG ACGCTAGGGC AGGGCATGAA
7501 AAAGCCCGTA GCGGGCTGCT ACGGGCGTCT GACGCGGTGG AAAGGGGGAG
7551 GGGATGTTGT CTACATGGCT CTGCTGTAGT GAGTGGGTTG CGCTCCGGCA
7601 GCGGTCCTGA TCAATCGTCA CCCTTTCTCG GTCCTTCAAC GTTCCTGACA
7651 ACGAGCCTCC TTTTCGCCAA TCCATCGACA ATCACCGCGA GTCCCTGCTC
7701 GAACGCTGCG TCCGGACCGG CTTCGTCGAA GGCGTCTATC GCGGCCCGCA
7751 ACAGCGGCGA GAGCGGAGCC TGTTCAACGG TGCCGCCGCG CTCGCCGGCA
7801 TCGCTGTCGC CGGCCTGCTC CTCAAGCACG GCCCCAACAG TGAAGTAGCT
7851 GATTGTCATC AGCGCATTGA CGGCGTCCCC GGCCGAAAAA CCCGCCTCGC
7901 AGAGGAAGCG AAGCTGCGCG TCGGCCGTTT CCATCTGCGG TGCGCCCGGT
7951 CGCGTGCCGG CATGGATGCG CGCGCCATCG CGGTAGGCGA GCAGCGCCTG
8001 CCTGAAGCTG CGGGCATTCC CGATCAGAAA TGAGCGCCAG TCGTCGTCGG
8051 CTCTCGGCAC CGAATGCGTA TGATTCTCCG CCAGCATGGC TTCGGCCAGT
8101 GCGTCGAGCA GCGCCCGCTT GTTCCTGAAG TGCCAGTAAA GCGCCGGCTG
8151 CTGAACCCCC AACCGTTCCG CCAGTTTGCG TGTCGTCAGA CCGTCTACGC
8201 CGACCTCGTT CAACAGGTCC AGGGCGGCAC GGATCACTGT ATTCGGCTGC
8251 AACTTTGTCA TGCTTGACAC TTTATCACTG ATAAACATAA TATGTCCACC
8301 AACTTATCAG TGATAAAGAA TCCGCGCGTT CAATCGGACC AGCGGAGGCT
8351 GGTCCGGAGG CCAGACATGA AACCCAACAT ACCCCTGATC GTAATTCTGA
8401 GCACTGTCGC GCTCGACGCT GTCGGCATCG GCCTGATTAT GCCGGTGCTG
8451 CCGGGCCTCC TGCGCGATCT GGTTCACTCG AACGACGTCA CCGCCCACTA
```

-continued

```
 8501 TGGCATTCTG CTGGCGCTGT ATGCGTTGGT GCAATTTGCC TGCGCACCTG
 8551 TGCTGGGCGC GCTGTCGGAT CGTTTCGGGC GGCGGCCAAT CTTGCTCGTC
 8601 TCGCTGGCCG GCGCCACTGT CGACTACGCC ATCATGGCGA CAGCGCCTTT
 8651 CCTTTGGGTT CTCTATATCG GGCGGATCGT GGCCGGCATC ACCGGGGCGA
 8701 CTGGGGCGGT AGCCGGCGCT TATATTGCCG ATATCACTGA TGGCGATGAG
 8751 CGCGCGCGGC ACTTCGGCTT CATGAGCGCC TGTTTCGGGT TCGGGATGGT
 8801 CGCGGGACCT GTGCTCGGTG GGCTGATGGG CGGTTTCTCC CCCCACGCTC
 8851 CGTTCTTCGC CGCGGCAGCC TTGAACGGCC TCAATTTCCT GACGGGCTGT
 8901 TTCCTTTTGC CGGAGTCGCA CAAAGGCGAA CGCCGGCCGT TACGCCGGGA
 8951 GGCTCTCAAC CCGCTCGCTT CGTTCCGGTG GGCCCGGGGC ATGACCGTCG
 9001 TCGCCGCCCT GATGGCGGTC TTCTTCATCA TGCAACTTGT CGGACAGGTG
 9051 CCGGCCGCGC TTTGGGTCAT TTTCGGCGAG GATCGCTTTC ACTGGGACGC
 9101 GACCACGATC GGCATTTCGC TTGCCGCATT TGGCATTCTG CATTCACTCG
 9151 CCCAGGCAAT GATCACCGGC CCTGTAGCCG CCCGGCTCGG CGAAAGGCGG
 9201 GCACTCATGC TCGGAATGAT TGCCGACGGC ACAGGCTACA TCCTGCTTGC
 9251 CTTCGCGACA CGGGGATGGA TGGCGTTCCC GATCATGGTC CTGCTTGCTT
 9301 CGGGTGGCAT CGGAATGCCG GCGCTGCAAG CAATGTTGTC CAGGCAGGTG
 9351 GATGAGGAAC GTCAGGGGCA GCTGCAAGGC TCACTGGCGG CGCTCACCAG
 9401 CCTGACCTCG ATCGTCGGAC CCCTCCTCTT CACGGCGATC TATGCGGCTT
 9451 CTATAACAAC GTGGAACGGG TGGGCATGGA TTGCAGGCGC TGCCCTCTAC
 9501 TTGCTCTGCC TGCCGGCGCT GCGTCGCGGG CTTTGGAGCG GCGCAGGGCA
 9551 ACGAGCCGAT CGCTGATCGT GGAAACGATA GGCCTATGCC ATGCGGGTCA
 9601 AGGCGACTTC CGGCAAGCTA TACGCGCCCT AGGAGTGCGG TTGGAACGTT
 9651 GGCCCAGCCA GATACTCCCG ATCACGAGCA GGACGCCGAT GATTTGAAGC
 9701 GCACTCAGCG TCTGATCCAA GAACAACCAT CCTAGCAACA CGGCGGTCCC
 9751 CGGGCTGAGA AAGCCCAGTA AGGAAACAAC TGTAGGTTCG AGTCGCGAGA
 9801 TCCCCCGGAA CCAAAGGAAG TAGGTTAAAC CCGCTCCGAT CAGGCCGAGC
 9851 CACGCCAGGC CGAGAACATT GGTTCCTGTA GGCATCGGGA TTGGCGGATC
 9901 AAACACTAAA GCTACTGGAA CGAGCAGAAG TCCTCCGGCC GCCAGTTGCC
 9951 AGGCGGTAAA GGTGAGCAGA GGCACGGGAG GTTGCCACTT GCGGGTCAGC
10001 ACGGTTCCGA ACGCCATGGA AACCGCCCCC GCCAGGCCCG CTGCGACGCC
10051 GACAGGATCT AGCGCTGCGT TTGGTGTCAA CACCAACAGC GCCACGCCCG
10101 CAGTTCCGCA AATAGCCCCC AGGACCGCCA TCAATCGTAT CGGGCTACCT
10151 AGCAGAGCGG CAGAGATGAA CACGACCATC AGCGGCTGCA CAGCGCCTAC

10201 CGTCGCCGCG ACCCCGCCCG GCAGGCGGTA GACCGAAATA AACAACAAGC
10251 TCCAGAATAG CGAAATATTA AGTGCGCCGA GGATGAAGAT GCGCATCCAC
10301 CAGATTCCCG TTGGAATCTG TCGGACGATC ATCACGAGCA ATAAACCCGC
10351 CGGCAACGCC CGCAGCAGCA TACCGGCGAC CCCTCGGCCT CGCTGTTCGG
10401 GCTCCACGAA AACGCCGGAC AGATGCGCCT TGTGAGCGTC CTTGGGGCCG
10451 TCCTCCTGTT TGAAGACCGA CAGCCCAATG ATCTCGCCGT CGATGTAGGC
10501 GCCGAATGCC ACGGCATCTC GCAACCGTTC AGCGAACGCC TCCATGGGCT
10551 TTTTCTCCTC GTGCTCGTAA ACGGACCCGA ACATCTCTGG AGCTTTCTTC
10601 AGGGCCGACA ATCGGATCTC GCGGAAATCC TGCACGTCGG CCGCTCCAAG
10651 CCGTCGAATC TGAGCCTTAA TCACAATTGT CAATTTTAAT CCTCTGTTTA
10701 TCGGCAGTTC GTAGAGCGCG CCGTGCGTCC CGAGCGATAC TGAGCGAAGC
10751 AAGTGCGTCG AGCAGTGCCC GCTTGTTCCT GAAATGCCAG TAAAGCGCTG
10801 GCTGCTGAAC CCCCAGCCGG AACTGACCCC ACAAGGCCCT AGCGTTTGCA
10851 ATGCACCAGG TCATCATTGA CCCAGGCGTG TTCCACCAGG CCGCTGCCTC
10901 GCAACTCTTC GCAGGCTTCG CCGACCTGCT CGCGCCACTT CTTCACGCGG
10951 GTGGAATCCG ATCCGCACAT GAGGCGGAAG GTTTCCAGCT TGAGCGGGTA
11001 CGGCTCCCGG TGCGAGCTGA AATAGTCGAA CATCCGTCGG GCCGTCGGCG
11051 ACAGCTTGCG GTACTTCTCC CATATGAATT TCGTGTAGTG GTCGCCAGCA
11101 AACAGCACGA CGATTTCCTC GTCGATCAGG ACCTGGCAAC GGGACGTTTT
11151 CTTGCCACGG TCCAGGACGC GGAAGCGGTG CAGCAGCGAC ACCGATTCCA
11201 GGTGCCCAAC GCGGTCGGAC GTGAAGCCCA TCGCCGTCGC CTGTAGGCGC
11251 GACAGGCATT CCTCGGCCTT CGTGTAATAC CGGCCATTGA TCGACCAGCC
11301 CAGGTCCTGG CAAAGCTCGT AGAACGTGAA GGTGATCGGC TCGCCGATAG
11351 GGGTGCGCTT CGCGTACTCC AACACCTGCT GCCACACCAG TTCGTCATCG
11401 TCGGCCCGCA GCTCGACGCC GGTGTAGGTG ATCTTCACGT CCTTGTTGAC
11451 GTGGAAAATG ACCTTGTTTT GCAGCGCCTC GCGCGGGATT TTCTTGTTGC
11501 GCGTGGTGAA CAGGGCAGAG CGGGCCGTGT CGTTTGGCAT CGCTCGCATC
11551 GTGTCCGGCC ACGGCGCAAT ATCGAACAAG GAAAGCTGCA TTTCCTTGAT
11601 CTGCTGCTTC GTGTGTTTCA GCAACGCGGC CTGCTTGGCC TCGCTGACCT
11651 GTTTTGCCAG GTCCTCGCCG GCGGTTTTTC GCTTCTTGGT CGTCATAGTT
11701 CCTCGCGTGT CGATGGTCAT CGACTTCGCC AAACCTGCCG CCTCCTGTTC
11751 GAGACGACGC GAACGCTCCA CGGCGGCCGA TGGCGCGGGC GGGCAGGGG
11801 GAGCCAGTTG CACGCTGTCG CGCTCGATCT TGGCCGTAGC TTGCTGGACC
11851 ATCGAGCCGA CGGACTGGAA GGTTTCGCGG GGCGCACGCA TGACGGTGCG
```

-continued

```
11901 GCTTGCGATG GTTTCGGCAT CCTCGGCGGA AAACCCCGCG TCGATCAGTT
11951 CTTGCCTGTA TGCCTTCCGG TCAAACGTCC GATTCATTCA CCCTCCTTGC
12001 GGGATTGCCC CGACTCACGC CGGGGCAATG TGCCCTTATT CCTGATTTGA
12051 CCCGCCTGGT GCCTTGGTGT CCAGATAATC CACCTTATCG GCAATGAAGT
12101 CGGTCCCGTA GACCGTCTGG CCGTCCTTCT CGTACTTGGT ATTCCGAATC
12151 TTGCCCTGCA CGAATACCAG CGACCCCTTG CCCAAATACT TGCCGTGGGC
12201 CTCGGCCTGA GAGCCAAAAC ACTTGATGCG GAAGAAGTCG GTGCGCTCCT
12251 GCTTGTCGCC GGTCGTGGCC GCGCCAACCT TTGCGATCCG CAAGCGCGCG
12301 GTCGCCATCT TCACGCTGGA ACAGTACGTC GAGGCGGGCA TCATGACCCG
12351 CGAGCAATAC GAGGTCATTA AAAGCGCCGT GATTGATGAT ATAGCGGCCC
12401 GGCTGCTCCT GGTTCTCGCG CACCGAAATG GGTGACTTCA CCCCGCGCTC
12451 TTTGATCGTG GCACCGATTT CCGCGATGCT CTCCGGGGAA AAGCCGGGGT
12501 TCTCGGCCGT CCGCGGCTGA TGCGGATCTT CGTCGATCAG GTCCAGGTCC
12551 AGCTCGATAG GGCCGGAACC GCCCTGAGAC GCCGCAGGAG CGTCCAGGAG
12601 GCTCGACAGG TCGCCGATGC TATCCAACCC CAGGCCGGAC GGCTGCGCCG
12651 CGCCTGCGGC TTCCTGAGCG GCCGCAGCGG TGTTTTTCTT GGTGGTCTTG
12701 GCTTGAGCCG CAGTCATTGG GAAATCTCCA TCTTCGTGAA CACGTAATCA
12751 GCCAGGGCGC GAACCTCTTT CGATGCCTTG CGCGCGGCCG TTTTCTTGAT
12801 CTTCCAGACC GGCACACCGG ATGCGAGGGC ATCGGCGATG CTGCTGCGCA
12851 GGCCAACGGT GGCCGGAATC ATCATCTTGG GGTACGCGGC CAGCAGCTCG
12901 GCTTGGTGGC GCGCGTGGCG CGGATTCCGC GCATCGACCT TGCTGGGCAC
12951 CATGCCAAGG AATTGCAGCT TGGCGTTCTT CTGGCGCACG TTCGCAATGG
13001 TCGTGACCAT CTTCTTGATG CCCTGGATGC TGTACGCCTC AAGCTCGATG
13051 GGGGACAGCA CATAGTCGGC CGCGAAGAGG GCGGCCGCCA GGCCGACGCC
13101 AAGGGTCGGG GCCGTGTCGA TCAGGCACAC GTCGAAGCCT TGGTTCGCCA
13151 GGGCCTTGAT GTTCGCCCCG AACAGCTCGC GGGCGTCGTC CAGCGACAGC
13201 CGTTCGGCGT TCGCCAGTAC CGGGTTGGAC TCGATGAGGG CGAGGCGCGC
13251 GGCCTGGCCG TCGCCGGCTG CGGGTGCGGT TTCGGTCCAG CCGCCGGCAG
13301 GGACAGCGCC GAACAGCTTG CTTGCATGCA GGCCGGTAGC AAAGTCCTTG
13351 AGCGTGTAGG ACGCATTGCC CTGGGGGTCC AGGTCGATCA CGGCAACCCG
13401 CAAGCCGCGC TCGAAAAAGT CGAAGGCAAG ATGCACAAGG GTCGAAGTCT
13451 TGCCGACGCC GCCTTTCTGG TTGGCCGTGA CCAAAGTTTT CATCGTTTGG
13501 TTTCCTGTTT TTTCTTGGCG TCCGCTTCCC ACTTCCGGAC GATGTACGCC
13551 TGATGTTCCG GCAGAACCGC CGTTACCCGC GCGTACCCCT CGGGCAAGTT
13601 CTTGTCCTCG AACGCGGCCC ACACGCGATG CACCGCTTGC GACACTGCGC
13651 CCCTGGTCAG TCCCAGCGAC GTTGCGAACG TCGCCTGTGG CTTCCCATCG
13701 ACTAAGACGC CCCGCGCTAT CTCGATGGTC TGCTGCCCCA CTTCCAGCCC
13751 CTGGATCGCC TCCTGGAACT GGCTTTCGGT AAGCCGTTTC TTCATGGATA
13801 ACACCCATAA TTTGCTCCGC GCCTTGGTTG AACATAGCGG TGACAGCCGC
13851 CAGCACATGA GAGAAGTTTA GCTAAACATT TCTCGCACGT CAACACCTTT
13901 AGCCGCTAAA ACTCGTCCTT GGCGTAACAA AACAAAAGCC CGGAAACCGG
13951 GCTTTCGTCT CTTGCCGCTT ATGGCTCTGC ACCCGGCTCC ATCACCAACA
14001 GGTCGCGCAC GCGCTTCACT CGGTTGCGGA TCGACACTGC CAGCCCAACA
14051 AAGCCGGTTG CCGCCGCCGC CAGGATCGCG CCGATGATGC CGGCCACACC
14101 GGCCATCGCC CACCAGGTCG CCGCCTTCCG GTTCCATTCC TGCTGGTACT
14151 GCTTCGCAAT GCTGGACCTC GGCTCACCAT AGGCTGACCG CTCGATGGCG
14201 TATGCCGCTT CTCCCCTTGG CGTAAAACCC AGCGCCGCAG GCGGCATTGC
14251 CATGCTGCCC GCCGCTTTCC CGACCACGAC GCGCGCACCA GGCTTGCGGT
14301 CCAGACCTTC GGCCACGGCG AGCTGCGCAA GGACATAATC AGCCGCCGAC
14351 TTGGCTCCAC GCGCCTCGAT CAGCTCTTGC ACTCGCGCGA AATCCTTGGC
14401 CTCCACGGCC GCCATGAATC GCGCACGCGG CGAAGGCTCC GCAGGGCCG
```

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 rtelated 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 Locht, C. (1992). Isolation and molecular characterization of a novel broad-host-range plasmid from *Bortadella 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. Cytochrone 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 cytochrone 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 tggtggtggt 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 gtggtggtgg 60

tggtgcatgg 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

aaggtatact atctagagtg gtggtggtgg tggtggtggt 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

aacccacgcc accagtaggc aggaggaaca cgtgtcgtcc 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 gaagggaaga 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 aaatagggaa 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccccct catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 900

tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 960

aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020

-continued

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 1080

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggaggatcg ccatgcacca ccaccaccac 2580

caccacgcct attccaatcc tacgtagaag ggaagatctt gagggccggc cctccgtcgc 2640

gggcggcacc cacgcccgca tcgattccaa ggttcagcca ttgagacggc tccgcttcgc 2700

gcgcaagcgc gggttgggcc gactgcaagc ggagagggaa gcatggcact gctcagcttc 2760

gagcgaaaat atcgcgtgcc gggggggcacg ctggtcggcg gaaacctgtt cgacttctgg 2820

gtcggccctt tctatgtcgg cttcttcggg gttgcgacgt ttttcttcgc ggccctgggt 2880

atcattctga ttgcctggag tgccgtactc cagggtacct ggaaccccca actcatctct 2940

gtctacccgc cggcccttga atatggcctg ggaggtgcac ccctcgcaaa aggcgggctg 3000

tggcagatca tcacgatctg cgccactggt gccttcgtca gctgggcgct gcgcgaagtc 3060

gaaatctgcc gtaagctggg catcgggtac cacatcccgt tcgccttcgc gttcgccatc 3120

ctggcctacc tgacgctggt gctgttccgc ccggtgatga tgggcgcctg gggctatgcc 3180

ttcccctacg ggatctggac gcacctcgac tgggtgtcga acacgggcta cacctacggc 3240

aacttccact acaaccctgc ccacatgatc gccatctcgt tcttcttcac gaacgcgctg 3300

-continued

gctctggcgc tgcacggcgc ccttgtgctc tccgcggcca accccgagaa gggcaaggaa 3360

atgcggacgc cggatcacga ggatacgttc ttccgcgatc tggtcggcta ctcgatcggg 3420

acgctcggca tccaccgcct cggcctgctg ctctcgctga gcgccgtctt cttcagcgcc 3480

ctctgcatga tcattaccgg caccatctgg ttcgatcagt gggtcgactg gtggcaatgg 3540

tgggtgaagc tgccgtggtg ggcgaacatc ccgggaggca tcaatggctg agtatcagaa 3600

catcttctcc caggtccagg tccgcggacc ggccgacctg gggatgaccg aagacgtcaa 3660

cctggccaac cgttcgggcg tcggtccctt ctcgaccctg ctcggctggt tcggcaacgc 3720

ccagctcggc ccgatctatc tcggctcgct cggcgtcctg tccctcttct cgggcctgat 3780

gtggttcttc accatcggga tctggttctg gtatcaggcg ggctggaacc cggccgtctt 3840

cctgcgcgac ctgttcttct tctcgctcga gccgccggca cccgaatacg gtctgtcctt 3900

cgcggctccg ctgaaggaag gcgggctgtg gctgatcgcg tcgttcttca tgttcgtcgc 3960

ggtctggtcc tggtggggcc gcacctatct ccgcgctcag gcgctgggca tgggcaagca 4020

caccgcctgg gcgttcctct cggccatctg gctgtggatg gtgctgggct tcatccgtcc 4080

gatcctcatg gggtcctggt cggaagcggt tccctacggc atcttctcgc acctcgactg 4140

gacgaacaac ttctcgctcg tccacggcaa cctgttctac aacccccttcc acggtctctc 4200

gatcgccttc ctctacgggt cggccctgct cttcgcgatg cacggtgcga ccatcctcgc 4260

ggtctcccgc ttcggcggcg agcgcgagct ggagcagatc gccgaccgcg ggacggcagc 4320

ggagcgggcc gccctcttct ggcgctggac catgggtttc aacgccacga tggaaggcat 4380

ccaccgctgg gccatctgga tggcggtcct cgtgaccctc accggcggca tcgggatcct 4440

gctctcgggc acggtcgtgg acaactggta cgtctggggc cagaaccacg gcatggcgcc 4500

gctgaactga ggagcgatca caatggctga caagaccatc ttcaacgatc acctcaacac 4560

caatccgaag accaaccttc gcctctgggt cgctttccag atgatgaagg gtgcgggctg 4620

ggctggcggc gtgttcttcg ggacgctcct tctcatcggg ttcttccggg tggtcgggcg 4680

gatgcttccg atccaggaga accaggctcc ggcgccgaac atcaccggcg ctctggagac 4740

cgggatcgag ctgatcaagc atctcgtctg agacaagtct cggggcaggg cggcgcgagg 4800

ccgcccgctc ctccaagtcc gggccatatc gccggcccgg gtccggggcg acaccacagc 4860

ccggttccct tcctgttggc gacagggacc tggtgccgtg tggaagaccg cacggcaccc 4920

ttttgacatt cacgggaggc tctgatgacc aatcccaccc cgcgacccga aaccccgctt 4980

ttggatcgcg tctgctgccc ggccgacatg aaggcgctga gtgacgccga actggagcgg 5040

ctggccgacg aagtgcgttc cgaggtcagt gatagggta gtttcttatt ttaggcagtt 5100

tatatgaaat taagacatgc agatgtcaca gtggatattg aactggtctc gaaagctcaa 5160

tatcccccaa agcacaagca caaacttcga catcatgcag aagcgtttcc cgaaccgcgt 5220

cttcgacgtg ggcatcgccg agcagcatgc cgtgaccttc gcggccggcc tcgccggggc 5280

cgggatgaag cccttctgcg cgatctattc ctcgttcctg caacggggtt acgaccagat 5340

cgcccatgac gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct 5400

cgtgggggcc gatggcgcga cccatgcggg ggccttcgac gttggcttca tcacttcgct 5460

gcccaacatg accgtgatgg ccgcggccga cgaggccgag ctcatccaca tgatcgccac 5520

cgccgtggcc ttcgacgagg gccccatcgc cttccgcttc ccgcggggcg agggggtggg 5580

-continued

```
cgtcgagatg cccgagcgcg ggacggtgct ggagcccggc cggggccgcg tggtgcgcga   5640

agggacggat gtcgcgatcc tctccttcgg cgcgcatctg cacgaggcct tgcaggcggc   5700

gaaacttctc gaggccgagg gggtgagcgt gaccgtggcc gacgcccgct tctcgcgccc   5760

gctcgacacg gggctcatcg accagctcgt gcgccatcac gcggcgctgg taacggtgga   5820

gcaggggccc atgggcggct tcggcgccca tgtcatgcac tatctcgcca attccggcgg   5880

cttcgacggg ggcctcgcgc tccgggtcat gacgctgccc gaccgcttca tcgagcaggc   5940

gagccccgag gacatgtatg ccgatgcggg gctgcgggcc gaggatatca agcttggcgt   6000

aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca   6060

tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat   6120

taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt   6180

aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctcg gtcttgcctt   6240

gctcgtcggt gatgtacttc accagctccg cgaagtcgct cttcttgatg gagcgcatgg   6300

ggacgtgctt ggcaatcacg cgcacccccc ggccgtttta gcggctaaaa aagtcatggc   6360

tctgccctcg ggcggaccac gcccatcatg accttgccaa gctcgtcctg cttctcttcg   6420

atcttcgcca gcagggcgag gatcgtggca tcaccgaacc gcgccgtgcg cgggtcgtcg   6480

gtgagccaga gtttcagcag gccgcccagg cggcccaggt cgccattgat gcgggccagc   6540

tcgcggacgt gctcatagtc cacgacgccc gtgattttgt agccctggcc gacggccagc   6600

aggtaggccg acaggctcat gccggccgcc gccgcctttt cctcaatcgc tcttcgttcg   6660

tctggaaggc agtacacctt gataggtggg ctgcccttcc tggttggctt ggtttcatca   6720

gccatccgct tgccctcatc tgttacgccg gcggtagccg gccagcctcg cagagcagga   6780

ttcccgttga gcaccgccag gtgcgaataa gggacagtga agaaggaaca cccgctcgcg   6840

ggtgggccta cttcacctat cctgcccggc tgacgccgtt ggatacacca aggaaagtct   6900

acacgaaccc tttggcaaaa tcctgtatat cgtgcgaaaa aggatggata taccgaaaaa   6960

atcgctataa tgaccccgaa gcagggttat gcagcggaaa agcgccacgc ttcccgaagg   7020

gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga   7080

gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact   7140

tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa   7200

cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc   7260

gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg   7320

ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgccagaa   7380

ggccgccaga gaggccgagc gcggccgtga ggcttggacg ctagggcagg gcatgaaaaa   7440

gcccgtagcg ggctgctacg ggcgtctgac gcggtggaaa gggggagggg atgttgtcta   7500

catggctctg ctgtagtgag tgggttgcgc tccggcagcg gtcctgatca atcgtcaccc   7560

tttctcggtc cttcaacgtt cctgacaacg agcctccttt tcgccaatcc atcgacaatc   7620

accgcgagtc cctgctcgaa cgctgcgtcc ggaccggctt cgtcgaaggc gtctatcgcg   7680

gcccgcaaca gcggcgagag cggagcctgt tcaacggtgc cgccgcgctc gccggcatcg   7740

ctgtcgccgg cctgctcctc aagcacggcc ccaacagtga agtagctgat tgtcatcagc   7800

gcattgacgg cgtccccggc cgaaaaaccc gcctcgcaga ggaagcgaag ctgcgcgtcg   7860
```

-continued

gccgtttcca tctgcggtgc gcccggtcgc gtgccggcat ggatgcgcgc gccatcgcgg 7920

taggcgagca gcgcctgcct gaagctgcgg gcattcccga tcagaaatga gcgccagtcg 7980

tcgtcggctc tcggcaccga atgcgtatga ttctccgcca gcatggcttc ggccagtgcg 8040

tcgagcagcg cccgcttgtt cctgaagtgc cagtaaagcg ccggctgctg aacccccaac 8100

cgttccgcca gtttgcgtgt cgtcagaccg tctacgccga cctcgttcaa caggtccagg 8160

gcggcacgga tcactgtatt cggctgcaac tttgtcatgc ttgacacttt atcactgata 8220

aacataatat gtccaccaac ttatcagtga taaagaatcc gcgcgttcaa tcggaccagc 8280

ggaggctggt ccggaggcca gacatgaaac ccaacatacc cctgatcgta attctgagca 8340

ctgtcgcgct cgacgctgtc ggcatcggcc tgattatgcc ggtgctgccg ggcctcctgc 8400

gcgatctggt tcactcgaac 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

gggcggtagc cggcgcttat attgccgata tcactgatgg cgatgagcgc gcgcggcact 8700

tcggcttcat gagcgcctgt ttcgggttcg ggatggtcgc gggacctgtg ctcggtgggc 8760

tgatgggcgg tttctccccc cacgctccgt tcttcgccgc ggcagccttg aacggcctca 8820

atttcctgac gggctgtttc cttttgccgg agtcgcacaa aggcgaacgc cggccgttac 8880

gccgggaggc tctcaacccg ctcgcttcgt tccggtgggc ccggggcatg accgtcgtcg 8940

ccgccctgat ggcggtcttc ttcatcatgc aacttgtcgg acaggtgccg gccgcgcttt 9000

gggtcatttt cggcgaggat cgctttcact gggacgcgac cacgatcggc atttcgcttg 9060

ccgcatttgg cattctgcat tcactcgccc aggcaatgat caccggccct gtagccgccc 9120

ggctcggcga aaggcgggca ctcatgctcg gaatgattgc cgacggcaca ggctacatcc 9180

tgcttgcctt cgcgacacgg ggatggatgg cgttcccgat catggtcctg cttgcttcgg 9240

gtggcatcgg aatgccggcg ctgcaagcaa tgttgtccag gcaggtggat gaggaacgtc 9300

aggggcagct gcaaggctca ctggcggcgc tcaccagcct gacctcgatc gtcggacccc 9360

tcctcttcac ggcgatctat gcggcttcta taacaacgtg gaacgggtgg gcatggattg 9420

caggcgctgc cctctacttg ctctgcctgc cggcgctgcg tcgcgggctt tggagcggcg 9480

cagggcaacg agccgatcgc tgatcgtgga aacgataggc ctatgccatg cgggtcaagg 9540

cgacttccgg caagctatac gcgccctagg agtgcggttg gaacgttggc ccagccagat 9600

actcccgatc acgagcagga cgccgatgat ttgaagcgca ctcagcgtct gatccaagaa 9660

caaccatcct agcaacacgg cggtccccgg 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 aggcccgctg cgacgccgac aggatctagc gctgcgtttg gtgtcaacac 10020

caacagcgcc acgcccgcag ttccgcaaat agcccccagg accgccatca atcgtatcgg 10080

gctacctagc agagcggcag agatgaacac gaccatcagc ggctgcacag cgcctaccgt 10140

-continued

cgccgcgacc ccgcccggca ggcggtagac cgaaataaac aacaagctcc agaatagcga 10200

aatattaagt gcgccgagga tgaagatgcg catccaccag attcccgttg gaatctgtcg 10260

gacgatcatc acgagcaata aacccgccgg caacgcccgc agcagcatac cggcgacccc 10320

tcggcctcgc tgttcgggct ccacgaaaac gccggacaga tgcgccttgt gagcgtcctt 10380

ggggccgtcc tcctgtttga agaccgacag cccaatgatc tcgccgtcga tgtaggcgcc 10440

gaatgccacg gcatctcgca accgttcagc gaacgcctcc atgggctttt tctcctcgtg 10500

ctcgtaaacg gacccgaaca tctctggagc tttcttcagg gccgacaatc ggatctcgcg 10560

gaaatcctgc acgtcggccg ctccaagccg tcgaatctga gccttaatca caattgtcaa 10620

ttttaatcct ctgtttatcg gcagttcgta gagcgcgccg tgcgtcccga gcgatactga 10680

gcgaagcaag tgcgtcgagc agtgcccgct tgttcctgaa atgccagtaa agcgctggct 10740

gctgaacccc cagccggaac tgaccccaca aggccctagc gtttgcaatg caccaggtca 10800

tcattgaccc aggcgtgttc caccaggccg ctgcctcgca actcttcgca ggcttcgccg 10860

acctgctcgc gccacttctt cacgcgggtg gaatccgatc cgcacatgag gcggaaggtt 10920

tccagcttga gcgggtacgg ctcccggtgc gagctgaaat agtcgaacat ccgtcgggcc 10980

gtcggcgaca gcttgcggta cttctcccat atgaatttcg tgtagtggtc gccagcaaac 11040

agcacgacga tttcctcgtc gatcaggacc tggcaacggg acgttttctt gccacggtcc 11100

aggacgcgga agcggtgcag cagcgacacc gattccaggt gcccaacgcg gtcggacgtg 11160

aagcccatcg ccgtcgcctg taggcgcgac aggcattcct cggccttcgt gtaataccgg 11220

ccattgatcg accagcccag gtcctggcaa agctcgtaga acgtgaaggt gatcggctcg 11280

ccgatagggg tgcgcttcgc gtactccaac acctgctgcc acaccagttc gtcatcgtcg 11340

gcccgcagct cgacgccggt gtaggtgatc ttcacgtcct tgttgacgtg gaaaatgacc 11400

ttgttttgca gcgcctcgcg cgggattttc ttgttgcgcg tggtgaacag ggcagagcgg 11460

gccgtgtcgt ttggcatcgc tcgcatcgtg tccggccacg gcgcaatatc gaacaaggaa 11520

agctgcattt ccttgatctg ctgcttcgtg tgtttcagca acgcggcctg cttggcctcg 11580

ctgacctgtt ttgccaggtc ctcgccggcg gtttttcgct tcttggtcgt catagttcct 11640

cgcgtgtcga tggtcatcga cttcgccaaa cctgccgcct cctgttcgag acgacgcgaa 11700

cgctccacgg cggccgatgg cgcgggcagg gcaggggggag ccagttgcac gctgtcgcgc 11760

tcgatcttgg ccgtagcttg ctggaccatc gagccgacgg actggaaggt ttcgcggggc 11820

gcacgcatga cggtgcggct tgcgatggtt tcggcatcct cggcggaaaa ccccgcgtcg 11880

atcagttctt gcctgtatgc cttccggtca aacgtccgat tcattcaccc tccttgcggg 11940

attgccccga ctcacgccgg ggcaatgtgc ccttattcct gatttgaccc gcctggtgcc 12000

ttggtgtcca gataatccac cttatcggca atgaagtcgg tcccgtagac cgtctggccg 12060

tccttctcgt acttggtatt ccgaatcttg ccctgcacga ataccagcga ccccttgccc 12120

aaatacttgc cgtgggcctc ggcctgagag ccaaaacact tgatgcggaa gaagtcggtg 12180

cgctcctgct tgtcgccggt cgtggccgcg ccaacctttg cgatccgcaa gcgcgcggtc 12240

gccatcttca cgctggaaca gtacgtcgag gcgggcatca tgacccgcga gcaatacgag 12300

gtcattaaaa gcgccgtgat tgatgatata gcggcccggc tgctcctggt tctcgcgcac 12360

cgaaatgggt gacttcaccc cgcgctcttt gatcgtggca ccgatttccg cgatgctctc 12420

-continued

cggggaaaag ccggggttgt cggccgtccg cggctgatgc ggatcttcgt cgatcaggtc 12480

caggtccagc tcgatagggc cggaaccgcc ctgagacgcc gcaggagcgt ccaggaggct 12540

cgacaggtcg ccgatgctat ccaaccccag gccggacggc tgcgccgcgc ctgcggcttc 12600

ctgagcggcc gcagcggtgt ttttcttggt ggtcttggct tgagccgcag tcattgggaa 12660

atctccatct tcgtgaacac gtaatcagcc agggcgcgaa cctctttcga tgccttgcgc 12720

gcggccgttt tcttgatctt ccagaccggc acaccggatg cgagggcatc ggcgatgctg 12780

ctgcgcaggc caacggtggc cggaatcatc atcttggggt acgcggccag cagctcggct 12840

tggtggcgcg cgtggcgcgg attccgcgca tcgaccttgc tgggcaccat gccaaggaat 12900

tgcagcttgg cgttcttctg gcgcacgttc gcaatggtcg tgaccatctt cttgatgccc 12960

tggatgctgt acgcctcaag ctcgatgggg gacagcacat agtcggccgc gaagagggcg 13020

gccgccaggc cgacgccaag ggtcggggcc gtgtcgatca ggcacacgtc gaagccttgg 13080

ttcgccaggg ccttgatgtt cgccccgaac agctcgcggg cgtcgtccag cgacagccgt 13140

tcggcgttcg ccagtaccgg gttggactcg atgagggcga ggcgcgcggc ctggccgtcg 13200

ccggctgcgg gtgcggtttc ggtccagccg ccggcaggga cagcgccgaa cagcttgctt 13260

gcatgcaggc cggtagcaaa gtccttgagc gtgtaggacg cattgccctg ggggtccagg 13320

tcgatcacgg caacccgcaa gccgcgctcg aaaaagtcga aggcaagatg cacaagggtc 13380

gaagtcttgc cgacgccgcc tttctggttg gccgtgacca aagttttcat cgtttggttt 13440

cctgtttttt cttggcgtcc gcttcccact tccggacgat gtacgcctga tgttccggca 13500

gaaccgccgt tacccgcgcg taccctcgg gcaagttctt gtcctcgaac gcggcccaca 13560

cgcgatgcac cgcttgcgac actgcgcccc tggtcagtcc cagcgacgtt gcgaacgtcg 13620

cctgtggctt cccatcgact aagacgcccc gcgctatctc gatggtctgc tgccccactt 13680

ccagcccctg gatcgcctcc tggaactggc tttcggtaag ccgtttcttc atggataaca 13740

cccataattt gctccgcgcc ttggttgaac atagcggtga cagccgccag cacatgagag 13800

aagtttagct aaacatttct cgcacgtcaa cacctttagc cgctaaaact cgtccttggc 13860

gtaacaaaac aaaagcccgg aaaccgggct ttcgtctctt gccgcttatg gctctgcacc 13920

cggctccatc accaacaggt cgcgcacgcg cttcactcgg ttgcggatcg acactgccag 13980

cccaacaaag ccggttgccg ccgccgccag gatcgcgccg atgatgccgg ccacaccggc 14040

catcgcccac caggtcgccg ccttccggtt ccattcctgc tggtactgct tcgcaatgct 14100

ggacctcggc tcaccatagg ctgaccgctc gatggcgtat gccgcttctc cccttggcgt 14160

aaaacccagc gccgcaggcg gcattgccat gctgcccgcc gctttcccga ccacgacgcg 14220

cgcaccaggc ttgcggtcca gaccttcggc cacggcgagc tgcgcaagga cataatcagc 14280

cgccgacttg gctccacgcg cctcgatcag ctcttgcact cgcgcgaaat ccttggcctc 14340

cacggccgcc atgaatcgcg cacgcggcga aggctccgca 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

```
ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc     60
tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc    120
ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg    180
gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg    240
agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt    300
aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc    360
gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa    420
tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc    480
ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag    540
gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga    600
ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc    660
gccgggtgag tcggcccctc aagtgtcaac gtccgccccct catctgtcag tgagggccaa    720
gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg    780
cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag    840
cccggtgagc gtcggaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg    900
tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag    960
aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc   1020
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca   1080
aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt   1140
ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc   1200
tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt   1260
tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt   1320
gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg   1380
acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc   1440
cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc   1500
gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg   1560
tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa   1620
acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc   1680
gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc   1740
cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg   1800
aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc   1860
cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa   1920
gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac   1980
agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa   2040
gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt   2100
gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga   2160
gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga   2220
gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac   2280
```

-continued

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggaggatcg ccatgcacca ccaccaccac 2580

caccacgcta gcgcgggcag atcttgaggc cggccctccg tcgcgggcgg cacccacgcc 2640

cgcatcgatt ccaaggttca gccattgaga cggctccgct tcgcgcgcaa gcgcgggttg 2700

ggccgactgc aagcggagag ggaagcatgg cactgctcag cttcgagcga aaatatcgcg 2760

tgccgggggg cacgctggtc ggcggaaacc tgttcgactt ctgggtcggc cctttctatg 2820

tcggcttctt cggggttgcg acgtttttct tcgcggccct gggtatcatt ctgattgcct 2880

ggagtgccgt actccagggt acctggaacc cccaactcat ctctgtctac ccgccggccc 2940

ttgaatatgg cctgggaggt gcacccctcg caaaaggcgg gctgtggcag atcatcacga 3000

tctgcgccac tggtgccttc gtcagctggg cgctgcgcga agtcgaaatc tgccgtaagc 3060

tgggcatcgg gtaccacatc ccgttcgcct tcgcgttcgc catcctggcc tacctgacgc 3120

tggtgctgtt ccgcccggtg atgatgggcg cctggggcta tgccttcccc tacgggatct 3180

ggacgcacct cgactgggtg tcgaacacgg gctacaccta cggcaacttc cactacaacc 3240

ctgcccacat gatcgccatc tcgttcttct tcacgaacgc gctggctctg gcgctgcacg 3300

gcgcccttgt gctctccgcg gccaaccccg agaagggcaa ggaaatgcgg acgccggatc 3360

acgaggatac gttcttccgc gatctggtcg gctactcgat cgggacgctc ggcatccacc 3420

gcctcggcct gctgctctcg ctgagcgccg tcttcttcag cgccctctgc atgatcatta 3480

ccggcaccat ctggttcgat cagtgggtcg actggtggca atggtgggtg aagctgccgt 3540

ggtgggcgaa catcccggga ggcatcaatg gctgagtatc agaacatctt ctcccaggtc 3600

caggtccgcg gaccggccga cctggggatg accgaagacg tcaacctggc caaccgttcg 3660

ggcgtcggtc ccttctcgac cctgctcggc tggttcggca acgcccagct cggcccgatc 3720

tatctcggct cgctcggcgt cctgtccctc ttctcgggcc tgatgtggtt cttcaccatc 3780

gggatctggt tctggtatca ggcgggctgg aacccggccg tcttcctgcg cgacctgttc 3840

ttcttctcgc tcgagccgcc ggcacccgaa tacggtctgt ccttcgcggc tccgctgaag 3900

gaaggcgggc tgtggctgat cgcgtcgttc ttcatgttcg tcgcggtctg gtcctggtgg 3960

ggccgcacct atctccgcgc tcaggcgctg ggcatgggca agcacaccgc ctgggcgttc 4020

ctctcggcca tctggctgtg gatggtgctg ggcttcatcc gtccgatcct catggggtcc 4080

tggtcggaag cggttcccta cggcatcttc tcgcacctcg actggacgaa caacttctcg 4140

ctcgtccacg gcaacctgtt ctacaacccc ttccacggtc tctcgatcgc cttcctctac 4200

gggtcggccc tgctcttcgc gatgcacggt gcgaccatcc tcgcggtctc ccgcttcggc 4260

ggcgagcgcg agctggagca gatcgccgac cgcgggacgg cagcggagcg ggccgccctc 4320

ttctggcgct ggaccatggg tttcaacgcc acgatggaag gcatccaccg ctgggccatc 4380

tggatggcgg tcctcgtgac cctcaccggc ggcatcggga tcctgctctc gggcacggtc 4440

gtggacaact ggtacgtctg gggccagaac cacggcatgg cgccgctgaa ctgaggagcg 4500

atcacaatgg ctgacaagac catcttcaac gatcacctca acaccaatcc gaagaccaac 4560

-continued

cttcgcctct gggtcgcttt ccagatgatg aagggtgcgg gctgggctgg cggcgtgttc 4620

ttcgggacgc tccttctcat cgggttcttc cgggtggtcg ggcggatgct tccgatccag 4680

gagaaccagg ctccggcgcc gaacatcacc ggcgctctgg agaccgggat cgagctgatc 4740

aagcatctcg tctgagacaa gtctcggggc agggcggcgc gaggccgccc gctcctccaa 4800

gtccgggcca tatcgccggc ccgggtccgg ggcgacacca cagcccggtt cccttcctgt 4860

tggcgacagg gacctggtgc cgtgtggaag accgcacggc acccttttga cattcacggg 4920

aggctctgat gaccaatccc accccgcgac ccgaaacccc gcttttggat cgcgtctgct 4980

gcccggccga catgaaggcg ctgagtgacg ccgaactgga gcggctggcc gacgaagtgc 5040

gttccgaggt cagtgatagg ggtagtttct tattttaggc agtttatatg aaattaagac 5100

atgcagatgt cacagtggat attgaactgg tctcgaaagc tcaatatccc ccaaagcaca 5160

agcacaaact tcgacatcat gcagaagcgt ttcccgaacc gcgtcttcga cgtgggcatc 5220

gccgagcagc atgccgtgac cttcgcggcc ggcctcgccg gggccgggat gaagcccttc 5280

tgcgcgatct attcctcgtt cctgcaacgg ggttacgacc agatcgccca tgacgtggcg 5340

ctgcagaacc ttcccgtccg cttcgtgatc gaccgggcgg ggctcgtggg ggccgatggc 5400

gcgacccatg cgggggcctt cgacgttggc ttcatcactt cgctgcccaa catgaccgtg 5460

atggccgcgg ccgacgaggc cgagctcatc cacatgatcg ccaccgccgt ggccttcgac 5520

gagggcccca tcgccttccg cttcccgcgg ggcgaggggg tgggcgtcga gatgcccgag 5580

cgcgggacgg tgctggagcc cggccggggc cgcgtggtgc gcgaagggac ggatgtcgcg 5640

atcctctcct tcggcgcgca tctgcacgag gccttgcagg cggcgaaact tctcgaggcc 5700

gaggggggtga gcgtgaccgt ggccgacgcc cgcttctcgc gcccgctcga cacggggctc 5760

atcgaccagc tcgtgcgcca tcacgcggcg ctggtaacgg tggagcaggg ggccatgggc 5820

ggcttcggcg cccatgtcat gcactatctc gccaattccg gcggcttcga cgggggcctc 5880

gcgctccggg tcatgacgct gcccgaccgc ttcatcgagc aggcgagccc cgaggacatg 5940

tatgccgatg cggggctgcg ggccgaggat atcaagcttg gcgtaatcat ggtcatagct 6000

gtttcctgtg tgaaattgtt atccgctcac aattccacac aacatacgag ccggaagcat 6060

aaagtgtaaa gcctggggtg cctaatgagt gagctaactc acattaattg cgttgcgctc 6120

actgcccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa tcggccaacg 6180

cgcggggaga ggcggtttgc gtattgggcg ctcggtcttg ccttgctcgt cggtgatgta 6240

cttcaccagc tccgcgaagt cgctcttctt gatggagcgc atggggacgt gcttggcaat 6300

cacgcgcacc ccccggccgt tttagcggct aaaaaagtca tggctctgcc ctcgggcgga 6360

ccacgcccat catgaccttg ccaagctcgt cctgcttctc ttcgatcttc gccagcaggg 6420

cgaggatcgt ggcatcaccg aaccgcgccg tgcgcgggtc gtcggtgagc cagagtttca 6480

gcaggccgcc caggcggccc aggtcgccat tgatgcgggc cagctcgcgg acgtgctcat 6540

agtccacgac gcccgtgatt ttgtagccct ggccgacggc cagcaggtag gccgacaggc 6600

tcatgccggc cgccgccgcc ttttcctcaa tcgctcttcg ttcgtctgga aggcagtaca 6660

ccttgatagg tgggctgccc ttcctggttg gcttggtttc atcagccatc cgcttgccct 6720

catctgttac gccggcggta gccggccagc ctcgcagagc aggattcccg ttgagcaccg 6780

ccaggtgcga ataagggaca gtgaagaagg aacacccgct cgcgggtggg cctacttcac 6840

-continued

ctatcctgcc cggctgacgc cgttggatac accaaggaaa gtctacacga acccttggc 6900

aaaatcctgt atatcgtgcg aaaaaggatg gatataccga aaaaatcgct ataatgaccc 6960

cgaagcaggg ttatgcagcg gaaaagcgcc acgcttcccg aagggagaaa ggcggacagg 7020

tatccggtaa gcggcagggt cggaacagga gagcgcacga gggagcttcc agggggaaac 7080

gcctggtatc tttatagtcc tgtcgggttt cgccacctct gacttgagcg tcgatttttg 7140

tgatgctcgt caggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg 7200

ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcgttatc ccctgattct 7260

gtggataacc gtattaccgc ctttgagtga gctgataccg ctcgccgcag ccgaacgacc 7320

gagcgcagcg agtcagtgag cgaggaagcg gaagagcgcc agaaggccgc cagagaggcc 7380

gagcgcggcc gtgaggcttg gacgctaggg cagggcatga aaaagcccgt agcgggctgc 7440

tacgggcgtc tgacgcggtg gaaaggggga ggggatgttg tctacatggc tctgctgtag 7500

tgagtgggtt gcgctccggc agcggtcctg atcaatcgtc acccttctc ggtccttcaa 7560

cgttcctgac aacgagcctc cttttcgcca atccatcgac aatcaccgcg agtccctgct 7620

cgaacgctgc gtccggaccg gcttcgtcga aggcgtctat cgcggcccgc aacagcggcg 7680

agagcggagc ctgttcaacg gtgccgccgc gctcgccggc atcgctgtcg ccggcctgct 7740

cctcaagcac ggccccaaca gtgaagtagc tgattgtcat cagcgcattg acggcgtccc 7800

cggccgaaaa acccgcctcg cagaggaagc gaagctgcgc gtcggccgtt tccatctgcg 7860

gtgcgcccgg tcgcgtgccg gcatggatgc gcgcgccatc gcggtaggcg agcagcgcct 7920

gcctgaagct gcgggcattc ccgatcagaa atgagcgcca gtcgtcgtcg gctctcggca 7980

ccgaatgcgt atgattctcc gccagcatgg cttcggccag tgcgtcgagc agcgcccgct 8040

tgttcctgaa gtgccagtaa agcgccggct gctgaacccc caaccgttcc gccagtttgc 8100

gtgtcgtcag accgtctacg ccgacctcgt tcaacaggtc cagggcggca cggatcactg 8160

tattcggctg caactttgtc atgcttgaca ctttatcact gataaacata atatgtccac 8220

caacttatca gtgataaaga atccgcgcgt tcaatcggac cagcggaggc tggtccggag 8280

gccagacatg aaacccaaca tacccctgat cgtaattctg agcactgtcg cgctcgacgc 8340

tgtcggcatc ggcctgatta tgccggtgct gccgggcctc ctgcgcgatc tggttcactc 8400

gaacgacgtc accgcccact atggcattct gctggcgctg tatgcgttgg tgcaatttgc 8460

ctgcgcacct gtgctgggcg cgctgtcgga tcgtttcggg cggcggccaa tcttgctcgt 8520

ctcgctggcc ggcgccactg tcgactacgc catcatggcg acagcgcctt tcctttgggt 8580

tctctatatc gggcggatcg tggccggcat caccggggcg actggggcgg tagccggcgc 8640

ttatattgcc gatatcactg atggcgatga gcgcgcgcgg cacttcggct tcatgagcgc 8700

ctgtttcggg ttcgggatgg tcgcgggacc tgtgctcggt gggctgatgg gcggtttctc 8760

cccccacgct ccgttcttcg ccgcggcagc cttgaacggc ctcaatttcc tgacgggctg 8820

tttccttttg ccggagtcgc acaaaggcga acgccggccg ttacgccggg aggctctcaa 8880

cccgctcgct tcgttccggt gggcccgggg catgaccgtc gtcgccgccc tgatggcggt 8940

cttcttcatc atgcaacttg tcggacaggt gccggccgcg ctttgggtca ttttcggcga 9000

ggatcgcttt cactgggacg cgaccacgat cggcatttcg cttgccgcat ttggcattct 9060

gcattcactc gcccaggcaa tgatcaccgg ccctgtagcc gcccggctcg gcgaaaggcg 9120

-continued

ggcactcatg ctcggaatga ttgccgacgg cacaggctac atcctgcttg ccttcgcgac 9180

acggggatgg atggcgttcc cgatcatggt cctgcttgct tcgggtggca tcggaatgcc 9240

ggcgctgcaa gcaatgttgt ccaggcaggt ggatgaggaa cgtcaggggc agctgcaagg 9300

ctcactggcg gcgctcacca gcctgacctc gatcgtcgga cccctcctct tcacggcgat 9360

ctatgcggct tctataacaa cgtggaacgg gtgggcatgg attgcaggcg ctgccctcta 9420

cttgctctgc ctgccggcgc tgcgtcgcgg gctttggagc ggcgcagggc aacgagccga 9480

tcgctgatcg tggaaacgat aggcctatgc catgcgggtc aaggcgactt ccggcaagct 9540

atacgcgccc taggagtgcg gttggaacgt tggcccagcc agatactccc gatcacgagc 9600

aggacgccga tgatttgaag cgcactcagc gtctgatcca agaacaacca tcctagcaac 9660

acggcggtcc ccgggctgag aaagcccagt aaggaaacaa ctgtaggttc gagtcgcgag 9720

atcccccgga accaaaggaa gtaggttaaa cccgctccga 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 tagcgctgcg tttggtgtca acaccaacag cgccacgccc 10020

gcagttccgc aaatagcccc caggaccgcc atcaatcgta tcgggctacc tagcagagcg 10080

gcagagatga acacgaccat cagcggctgc acagcgccta ccgtcgccgc gaccccgccc 10140

ggcaggcggt agaccgaaat aaacaacaag ctccagaata gcgaaatatt aagtgcgccg 10200

aggatgaaga tgcgcatcca ccagattccc gttggaatct gtcggacgat catcacgagc 10260

aataaacccg ccggcaacgc ccgcagcagc ataccggcga cccctcggcc tcgctgttcg 10320

ggctccacga aaacgccgga cagatgcgcc ttgtgagcgt ccttggggcc gtcctcctgt 10380

ttgaagaccg acagcccaat gatctcgccg tcgatgtagg cgccgaatgc cacggcatct 10440

cgcaaccgtt cagcgaacgc ctccatgggc tttttctcct cgtgctcgta aacggacccg 10500

aacatctctg gagctttctt cagggccgac aatcggatct cgcggaaatc ctgcacgtcg 10560

gccgctccaa gccgtcgaat ctgagcctta atcacaattg tcaattttaa tcctctgttt 10620

atcggcagtt cgtagagcgc gccgtgcgtc ccgagcgata ctgagcgaag caagtgcgtc 10680

gagcagtgcc cgcttgttcc tgaaatgcca gtaaagcgct ggctgctgaa cccccagccg 10740

gaactgaccc cacaaggccc tagcgtttgc aatgcaccag gtcatcattg acccaggcgt 10800

gttccaccag gccgctgcct cgcaactctt cgcaggcttc gccgacctgc tcgcgccact 10860

tcttcacgcg ggtggaatcc gatccgcaca tgaggcggaa ggtttccagc ttgagcgggt 10920

acggctcccg gtgcgagctg aaatagtcga acatccgtcg ggccgtcggc gacagcttgc 10980

ggtacttctc ccatatgaat ttcgtgtagt ggtcgccagc aaacagcacg acgatttcct 11040

cgtcgatcag gacctggcaa cgggacgttt tcttgccacg gtccaggacg cggaagcggt 11100

gcagcagcga caccgattcc aggtgcccaa cgcggtcgga cgtgaagccc atcgccgtcg 11160

cctgtaggcg cgacaggcat tcctcggcct tcgtgtaata ccggccattg atcgaccagc 11220

ccaggtcctg gcaaagctcg tagaacgtga aggtgatcgg ctcgccgata ggggtgcgct 11280

tcgcgtactc caacacctgc tgccacacca gttcgtcatc gtcggcccgc agctcgacgc 11340

cggtgtaggt gatcttcacg tccttgttga cgtggaaaat gaccttgttt tgcagcgcct 11400

-continued

cgcgcgggat tttcttgttg cgcgtggtga acagggcaga gcgggccgtg tcgtttggca 11460

tcgctcgcat cgtgtccggc cacggcgcaa tatcgaacaa ggaaagctgc atttccttga 11520

tctgctgctt cgtgtgtttc agcaacgcgg cctgcttggc ctcgctgacc tgttttgcca 11580

ggtcctcgcc ggcggttttt cgcttcttgg tcgtcatagt tcctcgcgtg tcgatggtca 11640

tcgacttcgc caaacctgcc gcctcctgtt cgagacgacg cgaacgctcc acggcggccg 11700

atggcgcggg cagggcaggg ggagccagtt gcacgctgtc gcgctcgatc ttggccgtag 11760

cttgctggac catcgagccg acggactgga aggtttcgcg gggcgcacgc atgacggtgc 11820

ggcttgcgat ggtttcggca tcctcggcgg aaaaccccgc gtcgatcagt tcttgcctgt 11880

atgccttccg gtcaaacgtc cgattcattc accctccttg cgggattgcc ccgactcacg 11940

ccggggcaat gtgcccttat tcctgatttg acccgcctgg tgccttggtg tccagataat 12000

ccaccttatc ggcaatgaag tcggtcccgt agaccgtctg gccgtccttc tcgtacttgg 12060

tattccgaat cttgccctgc acgaatacca gcgacccctt gcccaaatac ttgccgtggg 12120

cctcggcctg agagccaaaa cacttgatgc ggaagaagtc ggtgcgctcc tgcttgtcgc 12180

cggtcgtggc cgcgccaacc tttgcgatcc gcaagcgcgc ggtcgccatc ttcacgctgg 12240

aacagtacgt cgaggcgggc atcatgaccc gcgagcaata cgaggtcatt aaaagcgccg 12300

tgattgatga tatagcggcc cggctgctcc tggttctcgc gcaccgaaat gggtgacttc 12360

accccgcgct ctttgatcgt ggcaccgatt tccgcgatgc tctccgggga aaagccgggg 12420

ttgtcggccg tccgcggctg atgcggatct tcgtcgatca ggtccaggtc cagctcgata 12480

gggccggaac cgccctgaga cgccgcagga gcgtccagga ggctcgacag gtcgccgatg 12540

ctatccaacc ccaggccgga cggctgcgcc gcgcctgcgg cttcctgagc ggccgcagcg 12600

gtgtttttct tggtggtctt ggcttgagcc gcagtcattg ggaaatctcc atcttcgtga 12660

acacgtaatc agccagggcg cgaacctctt tcgatgcctt gcgcgcggcc gttttcttga 12720

tcttccagac cggcacaccg gatgcgaggg catcggcgat gctgctgcgc aggccaacgg 12780

tggccggaat catcatcttg gggtacgcgg ccagcagctc ggcttggtgg cgcgcgtggc 12840

gcggattccg cgcatcgacc ttgctgggca ccatgccaag gaattgcagc ttggcgttct 12900

tctggcgcac gttcgcaatg gtcgtgacca tcttcttgat gccctggatg ctgtacgcct 12960

caagctcgat gggggacagc acatagtcgg ccgcgaagag ggcggccgcc aggccgacgc 13020

caagggtcgg ggccgtgtcg atcaggcaca cgtcgaagcc ttggttcgcc agggccttga 13080

tgttcgcccc gaacagctcg cgggcgtcgt ccagcgacag ccgttcggcg ttcgccagta 13140

ccgggttgga ctcgatgagg gcgaggcgcg cggcctggcc gtcgccggct gcgggtgcgg 13200

tttcggtcca gccgccggca gggacagcgc cgaacagctt gcttgcatgc aggccggtag 13260

caaagtcctt gagcgtgtag gacgcattgc cctgggggtc caggtcgatc acggcaaccc 13320

gcaagccgcg ctcgaaaaag tcgaaggcaa gatgcacaag ggtcgaagtc ttgccgacgc 13380

cgcctttctg gttggccgtg accaaagttt tcatcgtttg gtttcctgtt ttttcttggc 13440

gtccgcttcc cacttccgga cgatgtacgc ctgatgttcc ggcagaaccg ccgttacccg 13500

cgcgtacccc tcgggcaagt tcttgtcctc gaacgcggcc cacacgcgat gcaccgcttg 13560

cgacactgcg cccctggtca gtcccagcga cgttgcgaac gtcgcctgtg gcttcccatc 13620

gactaagacg ccccgcgcta tctcgatggt ctgctgcccc acttccagcc cctggatcgc 13680

-continued

ctcctggaac tggctttcgg taagccgttt cttcatggat aacacccata atttgctccg 13740

cgccttggtt gaacatagcg gtgacagccg ccagcacatg agagaagttt agctaaacat 13800

ttctcgcacg tcaacacctt tagccgctaa aactcgtcct tggcgtaaca aaacaaaagc 13860

ccggaaaccg ggctttcgtc tcttgccgct tatggctctg cacccggctc catcaccaac 13920

aggtcgcgca cgcgcttcac tcggttgcgg atcgacactg ccagcccaac aaagccggtt 13980

gccgccgccg ccaggatcgc gccgatgatg ccggccacac cggccatcgc ccaccaggtc 14040

gccgccttcc ggttccattc ctgctggtac tgcttcgcaa tgctggacct cggctcacca 14100

taggctgacc gctcgatggc gtatgccgct tctcccccttg gcgtaaaacc cagcgccgca 14160

ggcggcattg ccatgctgcc cgccgctttc ccgaccacga cgcgcgcacc aggcttgcgg 14220

tccagacctt cggccacggc gagctgcgca aggacataat cagccgccga cttggctcca 14280

cgcgcctcga tcagctcttg cactcgcgcg aaatccttgg cctccacggc cgccatgaat 14340

cgcgcacgcg gcgaaggctc cgcagggccg 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 900

tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 960

aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 1080

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

-continued

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggaggatcg ccatgcacca ccaccaccac 2580

caccacgaga acctgtactt ccaatccttt acgtagaaat agggaagatc ttgagggccg 2640

gccctccgtc gcgggcggca cccacgcccg catcgattcc aaggttcagc cattgagacg 2700

gctccgcttc gcgcgcaagc gcgggttggg ccgactgcaa gcggagaggg aagcatggca 2760

ctgctcagct tcgagcgaaa atatcgcgtg ccgggggggca cgctggtcgg cggaaacctg 2820

ttcgacttct gggtcggccc tttctatgtc ggcttcttcg gggttgcgac gtttttcttc 2880

gcggccctgg gtatcattct gattgcctgg agtgccgtac tccagggtac ctggaacccc 2940

caactcatct ctgtctaccc gccggccctt gaatatggcc tgggaggtgc acccctcgca 3000

aaaggcgggc tgtggcagat catcacgatc tgcgccactg gtgccttcgt cagctgggcg 3060

ctgcgcgaag tcgaaatctg ccgtaagctg ggcatcgggt accacatccc gttcgccttc 3120

gcgttcgcca tcctggccta cctgacgctg gtgctgttcc gcccggtgat gatgggcgcc 3180

tggggctatg ccttccccta cgggatctgg acgcacctcg actgggtgtc gaacacgggc 3240

tacacctacg gcaacttcca ctacaaccct gcccacatga tcgccatctc gttcttcttc 3300

acgaacgcgc tggctctggc gctgcacggc gcccttgtgc tctccgcggc caaccccgag 3360

aagggcaagg aaatgcggac gccggatcac gaggatacgt tcttccgcga tctggtcggc 3420

tactcgatcg ggacgctcgg catccaccgc ctcggcctgc tgctctcgct gagcgccgtc 3480

-continued

ttcttcagcg ccctctgcat gatcattacc ggcaccatct ggttcgatca gtgggtcgac 3540

tggtggcaat ggtgggtgaa gctgccgtgg tgggcgaaca tcccgggagg catcaatggc 3600

tgagtatcag aacatcttct cccaggtcca ggtccgcgga ccggccgacc tggggatgac 3660

cgaagacgtc aacctggcca accgttcggg cgtcggtccc ttctcgaccc tgctcggctg 3720

gttcggcaac gcccagctcg gcccgatcta tctcggctcg ctcggcgtcc tgtccctctt 3780

ctcgggcctg atgtggttct tcaccatcgg gatctggttc tggtatcagg cgggctggaa 3840

cccggccgtc ttcctgcgcg acctgttctt cttctcgctc gagccgccgg cacccgaata 3900

cggtctgtcc ttcgcggctc cgctgaagga aggcgggctg tggctgatcg cgtcgttctt 3960

catgttcgtc gcggtctggt cctggtgggg ccgcacctat ctccgcgctc aggcgctggg 4020

catgggcaag cacaccgcct gggcgttcct ctcggccatc tggctgtgga tggtgctggg 4080

cttcatccgt ccgatcctca tggggtcctg gtcggaagcg gttccctacg gcatcttctc 4140

gcacctcgac tggacgaaca acttctcgct cgtccacggc aacctgttct acaaccccctt 4200

ccacggtctc tcgatcgcct tcctctacgg gtcggccctg ctcttcgcga tgcacggtgc 4260

gaccatcctc gcggtctccc gcttcggcgg cgagcgcgag ctggagcaga tcgccgaccg 4320

cgggacggca gcggagcggg ccgccctctt ctggcgctgg accatgggtt tcaacgccac 4380

gatggaaggc atccaccgct gggccatctg gatggcggtc ctcgtgaccc tcaccggcgg 4440

catcgggatc ctgctctcgg gcacggtcgt ggacaactgg tacgtctggg gccagaacca 4500

cggcatggcg ccgctgaact gaggagcgat cacaatggct gacaagacca tcttcaacga 4560

tcacctcaac accaatccga agaccaacct tcgcctctgg gtcgctttcc agatgatgaa 4620

gggtgcgggc tgggctggcg gcgtgttctt cgggacgctc cttctcatcg ggttcttccg 4680

ggtggtcggg cggatgcttc cgatccagga gaaccaggct ccggcgccga acatcaccgg 4740

cgctctggag accgggatcg agctgatcaa gcatctcgtc tgagacaagt ctcggggcag 4800

ggcggcgcga ggccgcccgc tcctccaagt ccgggccata tcgccggccc gggtccgggg 4860

cgacaccaca gcccggttcc cttcctgttg gcgacaggga cctggtgccg tgtggaagac 4920

cgcacggcac ccttttgaca ttcacgggag gctctgatga ccaatcccac cccgcgaccc 4980

gaaaccccgc ttttggatcg cgtctgctgc ccggccgaca tgaaggcgct gagtgacgcc 5040

gaactggagc ggctggccga cgaagtgcgt tccgaggtca gtgatagggg tagtttctta 5100

ttttaggcag tttatatgaa attaagacat gcagatgtca cagtggatat tgaactggtc 5160

tcgaaagctc aatatccccc aaagcacaag cacaaacttc gacatcatgc agaagcgttt 5220

cccgaaccgc gtcttcgacg tgggcatcgc cgagcagcat gccgtgacct tcgcggccgg 5280

cctcgccggg gccgggatga agcccttctg cgcgatctat tcctcgttcc tgcaacgggg 5340

ttacgaccag atcgcccatg acgtggcgct gcagaacctt cccgtccgct tcgtgatcga 5400

ccgggcgggg ctcgtggggg ccgatggcgc gacccatgcg ggggccttcg acgttggctt 5460

catcacttcg ctgcccaaca tgaccgtgat ggccgcggcc gacgaggccg agctcatcca 5520

catgatcgcc accgccgtgg ccttcgacga gggccccatc gccttccgct tcccgcgggg 5580

cgaggggtg ggcgtcgaga tgcccgagcg cgggacggtg ctggagcccg gccggggccg 5640

cgtggtgcgc gaagggacgg atgtcgcgat cctctccttc ggcgcgcatc tgcacgaggc 5700

cttgcaggcg gcgaaacttc tcgaggccga gggggtgagc gtgaccgtgg ccgacgcccg 5760

-continued

cttctcgcgc ccgctcgaca cggggctcat cgaccagctc gtgcgccatc acgcggcgct 5820

ggtaacggtg gagcaggggg ccatgggcgg cttcggcgcc catgtcatgc actatctcgc 5880

caattccggc ggcttcgacg ggggcctcgc gctccgggtc atgacgctgc ccgaccgctt 5940

catcgagcag gcgagccccg aggacatgta tgccgatgcg gggctgcggg ccgaggatat 6000

caagcttggc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacaa 6060

ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga 6120

gctaactcac attaattgcg ttgcgctcac tgcccgcttt ccagtcggga aacctgtcgt 6180

gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attgggcgct 6240

cggtcttgcc ttgctcgtcg gtgatgtact tcaccagctc cgcgaagtcg ctcttcttga 6300

tggagcgcat ggggacgtgc ttggcaatca cgcgcacccc ccggccgttt tagcggctaa 6360

aaaagtcatg gctctgccct cgggcggacc acgcccatca tgaccttgcc aagctcgtcc 6420

tgcttctctt cgatcttcgc cagcagggcg aggatcgtgg catcaccgaa ccgcgccgtg 6480

cgcgggtcgt cggtgagcca gagtttcagc aggccgccca ggcggcccag gtcgccattg 6540

atgcgggcca gctcgcggac gtgctcatag tccacgacgc ccgtgatttt gtagccctgg 6600

ccgacggcca gcaggtaggc cgacaggctc atgccggccg ccgccgcctt ttcctcaatc 6660

gctcttcgtt cgtctggaag gcagtacacc ttgataggtg ggctgccctt cctggttggc 6720

ttggtttcat cagccatccg cttgccctca tctgttacgc cggcggtagc cggccagcct 6780

cgcagagcag gattcccgtt gagcaccgcc aggtgcgaat aagggacagt gaagaaggaa 6840

cacccgctcg cgggtgggcc tacttcacct atcctgcccg gctgacgccg ttggatacac 6900

caaggaaagt ctacacgaac cctttggcaa aatcctgtat atcgtgcgaa aaaggatgga 6960

tataccgaaa aaatcgctat aatgaccccg aagcagggtt atgcagcgga aaagcgccac 7020

gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg gaacaggaga 7080

gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg 7140

ccacctctga cttgagcgtc gatttttgtg atgctcgtca gggggggcgga gcctatggaa 7200

aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt ttgctcacat 7260

gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct ttgagtgagc 7320

tgataccgct cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga 7380

agagcgccag aaggccgcca gagaggccga gcgcggccgt gaggcttgga cgctagggca 7440

gggcatgaaa aagcccgtag cgggctgcta cgggcgtctg acgcggtgga aaggggggagg 7500

ggatgttgtc tacatggctc tgctgtagtg agtgggttgc gctccggcag cggtcctgat 7560

caatcgtcac cctttctcgg tccttcaacg ttcctgacaa cgagcctcct tttcgccaat 7620

ccatcgacaa tcaccgcgag tccctgctcg aacgctgcgt ccggaccggc ttcgtcgaag 7680

gcgtctatcg cggcccgcaa cagcggcgag agcggagcct gttcaacggt gccgccgcgc 7740

tcgccggcat cgctgtcgcc ggcctgctcc tcaagcacgg ccccaacagt gaagtagctg 7800

attgtcatca gcgcattgac ggcgtccccg gccgaaaaac ccgcctcgca gaggaagcga 7860

agctgcgcgt cggccgtttc catctgcggt gcgcccggtc gcgtgccggc atggatgcgc 7920

gcgccatcgc ggtaggcgag cagcgcctgc ctgaagctgc gggcattccc gatcagaaat 7980

gagcgccagt cgtcgtcggc tctcggcacc gaatgcgtat gattctccgc cagcatggct 8040

-continued

tcggccagtg cgtcgagcag cgcccgcttg ttcctgaagt gccagtaaag cgccggctgc 8100

tgaaccccca accgttccgc cagtttgcgt gtcgtcagac cgtctacgcc gacctcgttc 8160

aacaggtcca gggcggcacg gatcactgta ttcggctgca actttgtcat gcttgacact 8220

ttatcactga taaacataat atgtccacca acttatcagt gataaagaat ccgcgcgttc 8280

aatcggacca gcggaggctg gtccggaggc cagacatgaa acccaacata cccctgatcg 8340

taattctgag cactgtcgcg ctcgacgctg tcggcatcgg cctgattatg ccggtgctgc 8400

cgggcctcct gcgcgatctg gttcactcga acgacgtcac cgcccactat ggcattctgc 8460

tggcgctgta tgcgttggtg caatttgcct gcgcacctgt gctgggcgcg ctgtcggatc 8520

gtttcgggcg gcggccaatc ttgctcgtct cgctggccgg cgccactgtc gactacgcca 8580

tcatggcgac agcgcctttc ctttgggttc tctatatcgg gcggatcgtg gccggcatca 8640

ccggggcgac tggggcggta gccggcgctt atattgccga tatcactgat ggcgatgagc 8700

gcgcgcggca cttcggcttc atgagcgcct gtttcgggtt cgggatggtc gcgggacctg 8760

tgctcggtgg gctgatgggc ggtttctccc cccacgctcc gttcttcgcc gcggcagcct 8820

tgaacggcct caatttcctg acgggctgtt tccttttgcc ggagtcgcac aaaggcgaac 8880

gccggccgtt acgccgggag gctctcaacc cgctcgcttc gttccggtgg gcccggggca 8940

tgaccgtcgt cgccgccctg atggcggtct tcttcatcat gcaacttgtc ggacaggtgc 9000

cggccgcgct ttgggtcatt ttcggcgagg atcgctttca ctgggacgcg accacgatcg 9060

gcatttcgct tgccgcattt ggcattctgc attcactcgc ccaggcaatg atcaccggcc 9120

ctgtagccgc ccggctcggc gaaaggcggg cactcatgct cggaatgatt gccgacggca 9180

caggctacat cctgcttgcc ttcgcgacac ggggatggat ggcgttcccg atcatggtcc 9240

tgcttgcttc gggtggcatc ggaatgccgg cgctgcaagc aatgttgtcc aggcaggtgg 9300

atgaggaacg tcaggggcag ctgcaaggct cactggcggc gctcaccagc ctgacctcga 9360

tcgtcggacc cctcctcttc acggcgatct atgcggcttc tataacaacg tggaacgggt 9420

gggcatggat tgcaggcgct gccctctact tgctctgcct gccggcgctg cgtcgcgggc 9480

tttggagcgg cgcagggcaa cgagccgatc gctgatcgtg gaaacgatag gcctatgcca 9540

tgcgggtcaa ggcgacttcc ggcaagctat acgcgcccta ggagtgcggt tggaacgttg 9600

gcccagccag atactcccga tcacgagcag gacgccgatg atttgaagcg cactcagcgt 9660

ctgatccaag aacaaccatc ctagcaacac ggcggtcccc gggctgagaa agcccagtaa 9720

ggaaacaact gtaggttcga gtcgcgagat cccccggaac caaaggaagt aggttaaacc 9780

cgctccgatc aggccgagcc acgccaggcc gagaacattg gttcctgtag gcatcgggat 9840

tggcggatca aacactaaag ctactggaac gagcagaagt cctccggccg ccagttgcca 9900

ggcggtaaag gtgagcagag gcacgggagg ttgccacttg cgggtcagca cggttccgaa 9960

cgccatggaa accgcccccg ccaggcccgc tgcgacgccg acaggatcta gcgctgcgtt 10020

tggtgtcaac accaacagcg ccacgcccgc agttccgcaa atagccccca ggaccgccat 10080

caatcgtatc gggctaccta gcagagcggc agagatgaac acgaccatca gcggctgcac 10140

agcgcctacc gtcgccgcga ccccgcccgg caggcggtag accgaaataa acaacaagct 10200

ccagaatagc gaaatattaa gtgcgccgag gatgaagatg cgcatccacc agattcccgt 10260

tggaatctgt cggacgatca tcacgagcaa taaacccgcc ggcaacgccc gcagcagcat 10320

-continued

accggcgacc cctcggcctc gctgttcggg ctccacgaaa acgccggaca gatgcgcctt 10380

gtgagcgtcc ttggggccgt cctcctgttt gaagaccgac agcccaatga tctcgccgtc 10440

gatgtaggcg ccgaatgcca cggcatctcg caaccgttca gcgaacgcct ccatgggctt 10500

tttctcctcg tgctcgtaaa cggacccgaa catctctgga gctttcttca gggccgacaa 10560

tcggatctcg cggaaatcct gcacgtcggc cgctccaagc cgtcgaatct gagccttaat 10620

cacaattgtc aattttaatc ctctgtttat cggcagttcg tagagcgcgc cgtgcgtccc 10680

gagcgatact gagcgaagca agtgcgtcga gcagtgcccg cttgttcctg aaatgccagt 10740

aaagcgctgg ctgctgaacc cccagccgga actgacccca caaggcccta gcgtttgcaa 10800

tgcaccaggt catcattgac ccaggcgtgt tccaccaggc cgctgcctcg caactcttcg 10860

caggcttcgc cgacctgctc gcgccacttc ttcacgcggg tggaatccga tccgcacatg 10920

aggcggaagg tttccagctt gagcgggtac ggctcccggt gcgagctgaa atagtcgaac 10980

atccgtcggg ccgtcggcga cagcttgcgg tacttctccc atatgaattt cgtgtagtgg 11040

tcgccagcaa acagcacgac gatttcctcg tcgatcagga cctggcaacg ggacgttttc 11100

ttgccacggt ccaggacgcg gaagcggtgc agcagcgaca ccgattccag gtgcccaacg 11160

cggtcggacg tgaagcccat cgccgtcgcc tgtaggcgcg acaggcattc ctcggccttc 11220

gtgtaatacc ggccattgat cgaccagccc aggtcctggc aaagctcgta gaacgtgaag 11280

gtgatcggct cgccgatagg ggtgcgcttc gcgtactcca acacctgctg ccacaccagt 11340

tcgtcatcgt cggcccgcag ctcgacgccg gtgtaggtga tcttcacgtc cttgttgacg 11400

tggaaaatga ccttgttttg cagcgcctcg cgcgggattt tcttgttgcg cgtggtgaac 11460

agggcagagc gggccgtgtc gtttggcatc gctcgcatcg tgtccggcca cggcgcaata 11520

tcgaacaagg aaagctgcat ttccttgatc tgctgcttcg tgtgtttcag caacgcggcc 11580

tgcttggcct cgctgacctg ttttgccagg tcctcgccgg cggtttttcg cttcttggtc 11640

gtcatagttc ctcgcgtgtc gatggtcatc gacttcgcca aacctgccgc ctcctgttcg 11700

agacgacgcg aacgctccac ggcggccgat ggcgcgggca gggcaggggg agccagttgc 11760

acgctgtcgc gctcgatctt ggccgtagct tgctggacca tcgagccgac ggactggaag 11820

gtttcgcggg gcgcacgcat gacggtgcgg cttgcgatgg tttcggcatc ctcggcggaa 11880

aaccccgcgt cgatcagttc ttgcctgtat gccttccggt caaacgtccg attcattcac 11940

cctccttgcg ggattgcccc gactcacgcc ggggcaatgt gcccttattc ctgatttgac 12000

ccgcctggtg ccttggtgtc cagataatcc accttatcgg caatgaagtc ggtcccgtag 12060

accgtctggc cgtccttctc gtacttggta ttccgaatct tgccctgcac gaataccagc 12120

gaccccttgc ccaaatactt gccgtgggcc tcggcctgag agccaaaaca cttgatgcgg 12180

aagaagtcgg tgcgctcctg cttgtcgccg gtcgtggccg cgccaacctt tgcgatccgc 12240

aagcgcgcgg tcgccatctt cacgctggaa cagtacgtcg aggcgggcat catgacccgc 12300

gagcaatacg aggtcattaa aagcgccgtg attgatgata tagcggcccg gctgctcctg 12360

gttctcgcgc accgaaatgg gtgacttcac cccgcgctct ttgatcgtgg caccgatttc 12420

cgcgatgctc tccggggaaa agccggggtt gtcggccgtc cgcggctgat gcggatcttc 12480

gtcgatcagg tccaggtcca gctcgatagg gccggaaccg ccctgagacg ccgcaggagc 12540

gtccaggagg ctcgacaggt cgccgatgct atccaacccc aggccggacg gctgcgccgc 12600

-continued

gcctgcggct tcctgagcgg ccgcagcggt gtttttcttg gtggtcttgg cttgagccgc 12660

agtcattggg aaatctccat cttcgtgaac acgtaatcag ccagggcgcg aacctctttc 12720

gatgccttgc gcgcggccgt tttcttgatc ttccagaccg gcacaccgga tgcgagggca 12780

tcggcgatgc tgctgcgcag gccaacggtg gccggaatca tcatcttggg gtacgcggcc 12840

agcagctcgg cttggtggcg cgcgtggcgc ggattccgcg catcgacctt gctgggcacc 12900

atgccaagga attgcagctt ggcgttcttc tggcgcacgt tcgcaatggt cgtgaccatc 12960

ttcttgatgc cctggatgct gtacgcctca agctcgatgg gggacagcac atagtcggcc 13020

gcgaagaggg cggccgccag gccgacgcca agggtcgggg ccgtgtcgat caggcacacg 13080

tcgaagcctt ggttcgccag ggccttgatg ttcgccccga acagctcgcg ggcgtcgtcc 13140

agcgacagcc gttcggcgtt cgccagtacc gggttggact cgatgagggc gaggcgcgcg 13200

gcctggccgt cgccggctgc gggtgcggtt tcggtccagc cgccggcagg gacagcgccg 13260

aacagcttgc ttgcatgcag gccggtagca aagtccttga gcgtgtagga cgcattgccc 13320

tgggggtcca ggtcgatcac ggcaacccgc aagccgcgct cgaaaaagtc gaaggcaaga 13380

tgcacaaggg tcgaagtctt gccgacgccg cctttctggt tggccgtgac caaagttttc 13440

atcgtttggt ttcctgtttt ttcttggcgt ccgcttccca cttccggacg atgtacgcct 13500

gatgttccgg cagaaccgcc gttacccgcg cgtacccctc gggcaagttc ttgtcctcga 13560

acgcggccca cacgcgatgc accgcttgcg acactgcgcc cctggtcagt cccagcgacg 13620

ttgcgaacgt cgcctgtggc ttcccatcga ctaagacgcc ccgcgctatc tcgatggtct 13680

gctgccccac ttccagcccc tggatcgcct cctggaactg gctttcggta agccgtttct 13740

tcatggataa cacccataat ttgctccgcg ccttggttga acatagcggt gacagccgcc 13800

agcacatgag agaagtttag ctaaacattt ctcgcacgtc aacaccttta gccgctaaaa 13860

ctcgtccttg gcgtaacaaa acaaaagccc ggaaaccggg ctttcgtctc ttgccgctta 13920

tggctctgca cccggctcca tcaccaacag gtcgcgcacg cgcttcactc ggttgcggat 13980

cgacactgcc agcccaacaa agccggttgc cgccgccgcc aggatcgcgc cgatgatgcc 14040

ggccacaccg gccatcgccc accaggtcgc cgccttccgg ttccattcct gctggtactg 14100

cttcgcaatg ctggacctcg gctcaccata ggctgaccgc tcgatggcgt atgccgcttc 14160

tcccccttggc gtaaaaccca gcgccgcagg cggcattgcc atgctgcccg ccgctttccc 14220

gaccacgacg cgcgcaccag gcttgcggtc cagaccttcg gccacggcga gctgcgcaag 14280

gacataatca gccgccgact tggctccacg cgcctcgatc agctcttgca ctcgcgcgaa 14340

atccttggcc tccacggccg ccatgaatcg cgcacgcggc gaaggctccg cagggccg 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

-continued

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggag gcaccccccc tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccccc catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

-continued

```
ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg   2520

tgccttaccg taacccacgc caccagcatg tggaggatcg ccatgcacca ccaccaccac   2580

caccacgaga acctgtactt ccagtccgct agctagggaa gatcttgagg ccggccctcc   2640

gtcgcgggcg gcacccacgc ccgcatcgat tccaaggttc agccattgag acggctccgc   2700

ttcgcgcgca agcgcgggtt gggccgactg caagcggaga gggaagcatg gcactgctca   2760

gcttcgagcg aaaatatcgc gtgccggggg gcacgctggt cggcggaaac ctgttcgact   2820

tctgggtcgg ccctttctat gtcggcttct tcggggttgc gacgtttttc ttcgcggccc   2880

tgggtatcat tctgattgcc tggagtgccg tactccaggg tacctggaac ccccaactca   2940

tctctgtcta cccgccggcc cttgaatatg gcctgggagg tgcacccctc gcaaaaggcg   3000

ggctgtggca gatcatcacg atctgcgcca ctggtgcctt cgtcagctgg gcgctgcgcg   3060

aagtcgaaat ctgccgtaag ctgggcatcg ggtaccacat cccgttcgcc ttcgcgttcg   3120

ccatcctggc ctacctgacg ctggtgctgt tccgcccggt gatgatgggc gcctggggct   3180

atgccttccc ctacgggatc tggacgcacc tcgactgggt gtcgaacacg ggctacacct   3240

acggcaactt ccactacaac cctgcccaca tgatcgccat ctcgttcttc ttcacgaacg   3300

cgctggctct ggcgctgcac ggcgcccttg tgctctccgc ggccaacccc gagaagggca   3360

aggaaatgcg gacgccggat cacgaggata cgttcttccg cgatctggtc ggctactcga   3420

tcgggacgct cggcatccac cgcctcggcc tgctgctctc gctgagcgcc gtcttcttca   3480

gcgccctctg catgatcatt accggcacca tctggttcga tcagtgggtc gactggtggc   3540

aatggtgggt gaagctgccg tggtgggcga acatcccggg aggcatcaat ggctgagtat   3600

cagaacatct tctcccaggt ccaggtccgc ggaccggccg acctggggat gaccgaagac   3660

gtcaacctgg ccaaccgttc gggcgtcggt cccttctcga ccctgctcgg ctggttcggc   3720

aacgcccagc tcggcccgat ctatctcggc tcgctcggcg tcctgtccct cttctcgggc   3780

ctgatgtggt tcttcaccat cgggatctgg ttctggtatc aggcgggctg gaacccggcc   3840

gtcttcctgc gcgacctgtt cttcttctcg ctcgagccgc cggcacccga atacggtctg   3900

tccttcgcgg ctccgctgaa ggaaggcggg ctgtggctga tcgcgtcgtt cttcatgttc   3960

gtcgcggtct ggtcctggtg gggccgcacc tatctccgcg ctcaggcgct gggcatgggc   4020

aagcacaccg cctgggcgtt cctctcggcc atctggctgt ggatggtgct gggcttcatc   4080

cgtccgatcc tcatggggtc ctggtcggaa gcggttccct acggcatctt ctcgcacctc   4140

gactggacga acaacttctc gctcgtccac ggcaacctgt tctacaaccc cttccacggt   4200

ctctcgatcg ccttcctcta cgggtcggcc ctgctcttcg cgatgcacgg tgcgaccatc   4260

ctcgcggtct cccgcttcgg cggcgagcgc gagctggagc agatcgccga ccgcgggacg   4320

gcagcggagc gggccgccct cttctggcgc tggaccatgg gtttcaacgc cacgatggaa   4380

ggcatccacc gctgggccat ctggatggcg gtcctcgtga ccctcaccgg cggcatcggg   4440

atcctgctct cgggcacggt cgtggacaac tggtacgtct ggggccagaa ccacggcatg   4500

gcgccgctga actgaggagc gatcacaatg gctgacaaga ccatcttcaa cgatcacctc   4560

aacaccaatc cgaagaccaa ccttcgcctc tgggtcgctt tccagatgat gaagggtgcg   4620

ggctgggctg gcggcgtgtt cttcgggacg ctccttctca tcgggttctt ccgggtggtc   4680

gggcggatgc ttccgatcca ggagaaccag gctccggcgc cgaacatcac cggcgctctg   4740
```

-continued

gagaccggga tcgagctgat caagcatctc gtctgagaca agtctcgggg cagggcggcg 4800

cgaggccgcc cgctcctcca agtccgggcc atatcgccgg cccgggtccg gggcgacacc 4860

acagcccggt tcccttcctg ttggcgacag ggacctggtg ccgtgtggaa gaccgcacgg 4920

cacccttttg acattcacgg gaggctctga tgaccaatcc caccccgcga cccgaaaccc 4980

cgcttttgga tcgcgtctgc tgcccggccg acatgaaggc gctgagtgac gccgaactgg 5040

agcggctggc cgacgaagtg cgttccgagg tcagtgatag gggtagtttc ttattttagg 5100

cagtttatat gaaattaaga catgcagatg tcacagtgga tattgaactg gtctcgaaag 5160

ctcaatatcc cccaaagcac aagcacaaac ttcgacatca tgcagaagcg tttcccgaac 5220

cgcgtcttcg acgtgggcat cgccgagcag catgccgtga ccttcgcggc cggcctcgcc 5280

ggggccggga tgaagccctt ctgcgcgatc tattcctcgt tcctgcaacg gggttacgac 5340

cagatcgccc atgacgtggc gctgcagaac cttcccgtcc gcttcgtgat cgaccgggcg 5400

gggctcgtgg gggccgatgg cgcgacccat gcgggggcct tcgacgttgg cttcatcact 5460

tcgctgccca acatgaccgt gatggccgcg gccgacgagg ccgagctcat ccacatgatc 5520

gccaccgccg tggccttcga cgagggcccc atcgccttcc gcttcccgcg gggcgagggg 5580

gtgggcgtcg agatgcccga gcgcgggacg gtgctggagc ccggccgggg ccgcgtggtg 5640

cgcgaaggga cggatgtcgc gatcctctcc ttcggcgcgc atctgcacga ggccttgcag 5700

gcggcgaaac ttctcgaggc cgagggggtg agcgtgaccg tggccgacgc ccgcttctcg 5760

cgcccgctcg acacggggct catcgaccag ctcgtgcgcc atcacgcggc gctggtaacg 5820

gtggagcagg gggccatggg cggcttcggc gcccatgtca tgcactatct cgccaattcc 5880

ggcggcttcg acgggggcct cgcgctccgg gtcatgacgc tgcccgaccg cttcatcgag 5940

caggcgagcc ccgaggacat gtatgccgat gcggggctgc gggccgagga tatcaagctt 6000

ggcgtaatca tggtcatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca 6060

caacatacga gccggaagca taaagtgtaa agcctggggt gcctaatgag tgagctaact 6120

cacattaatt gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt cgtgccagct 6180

gcattaatga atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcggtctt 6240

gccttgctcg tcggtgatgt acttcaccag ctccgcgaag tcgctcttct tgatggagcg 6300

catggggacg tgcttggcaa tcacgcgcac cccccggccg ttttagcggc taaaaaagtc 6360

atggctctgc cctcgggcgg accacgccca tcatgacctt gccaagctcg tcctgcttct 6420

cttcgatctt cgccagcagg gcgaggatcg tggcatcacc gaaccgcgcc gtgcgcgggt 6480

cgtcggtgag ccagagtttc agcaggccgc ccaggcggcc caggtcgcca ttgatgcggg 6540

ccagctcgcg gacgtgctca tagtccacga cgcccgtgat tttgtagccc tggccgacgg 6600

ccagcaggta ggccgacagg ctcatgccgg ccgccgccgc cttttcctca atcgctcttc 6660

gttcgtctgg aaggcagtac accttgatag gtgggctgcc cttcctggtt ggcttggttt 6720

catcagccat ccgcttgccc tcatctgtta cgccggcggt agccggccag cctcgcagag 6780

caggattccc gttgagcacc gccaggtgcg aataagggac agtgaagaag gaacacccgc 6840

tcgcgggtgg gcctacttca cctatcctgc ccggctgacg ccgttggata caccaaggaa 6900

agtctacacg aacccttttgg caaaatcctg tatatcgtgc gaaaaaggat ggatataccg 6960

aaaaaatcgc tataatgacc ccgaagcagg gttatgcagc ggaaaagcgc cacgcttccc 7020

-continued

gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg 7080

agggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc 7140

tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc 7200

agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt 7260

cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg agctgatacc 7320

gctcgccgca gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc 7380

cagaaggccg ccagagaggc cgagcgcggc cgtgaggctt ggacgctagg gcagggcatg 7440

aaaaagcccg tagcgggctg ctacgggcgt ctgacgcggt ggaaaggggg aggggatgtt 7500

gtctacatgg ctctgctgta gtgagtgggt tgcgctccgg cagcggtcct gatcaatcgt 7560

cacccttttct cggtccttca acgttcctga caacgagcct ccttttcgcc aatccatcga 7620

caatcaccgc gagtccctgc tcgaacgctg cgtccggacc ggcttcgtcg aaggcgtcta 7680

tcgcggcccg caacagcggc gagagcggag cctgttcaac ggtgccgccg cgctcgccgg 7740

catcgctgtc gccggcctgc tcctcaagca cggccccaac agtgaagtag ctgattgtca 7800

tcagcgcatt gacggcgtcc ccggccgaaa aacccgcctc gcagaggaag cgaagctgcg 7860

cgtcggccgt ttccatctgc ggtgcgcccg gtcgcgtgcc ggcatggatg cgcgcgccat 7920

cgcggtaggc gagcagcgcc tgcctgaagc tgcgggcatt cccgatcaga aatgagcgcc 7980

agtcgtcgtc ggctctcggc accgaatgcg tatgattctc cgccagcatg gcttcggcca 8040

gtgcgtcgag cagcgcccgc ttgttcctga agtgccagta aagcgccggc tgctgaaccc 8100

ccaaccgttc cgccagtttg cgtgtcgtca gaccgtctac gccgacctcg ttcaacaggt 8160

ccagggcggc acggatcact gtattcggct gcaactttgt catgcttgac actttatcac 8220

tgataaacat aatatgtcca ccaacttatc agtgataaag aatccgcgcg ttcaatcgga 8280

ccagcggagg ctggtccgga ggccagacat gaaacccaac ataccctga tcgtaattct 8340

gagcactgtc gcgctcgacg ctgtcggcat cggcctgatt atgccggtgc tgccgggcct 8400

cctgcgcgat ctggttcact cgaacgacgt caccgcccac tatggcattc tgctggcgct 8460

gtatgcgttg gtgcaatttg cctgcgcacc tgtgctgggc gcgctgtcgg atcgtttcgg 8520

gcggcggcca atcttgctcg tctcgctggc cggcgccact gtcgactacg ccatcatggc 8580

gacagcgcct ttcctttggg ttctctatat cgggcggatc gtggccggca tcaccggggc 8640

gactggggcg gtagccggcg cttatattgc cgatatcact gatggcgatg agcgcgcgcg 8700

gcacttcggc ttcatgagcg cctgtttcgg gttcgggatg gtcgcgggac ctgtgctcgg 8760

tgggctgatg ggcggtttct ccccccacgc tccgttcttc gccgcggcag ccttgaacgg 8820

cctcaatttc ctgacgggct gtttcctttt gccggagtcg cacaaaggcg aacgccggcc 8880

gttacgccgg gaggctctca acccgctcgc ttcgttccgg tgggcccggg gcatgaccgt 8940

cgtcgccgcc ctgatggcgg tcttcttcat catgcaactt gtcggacagg tgccggccgc 9000

gctttgggtc attttcggcg aggatcgctt tcactgggac gcgaccacga tcggcatttc 9060

gcttgccgca tttggcattc tgcattcact cgcccaggca atgatcaccg gccctgtagc 9120

cgcccggctc ggcgaaaggc gggcactcat gctcggaatg attgccgacg gcacaggcta 9180

catcctgctt gccttcgcga cacggggatg gatggcgttc ccgatcatgg tcctgcttgc 9240

ttcgggtggc atcggaatgc cggcgctgca agcaatgttg tccaggcagg tggatgagga 9300

-continued

acgtcagggg cagctgcaag gctcactggc ggcgctcacc agcctgacct cgatcgtcgg 9360

acccctcctc ttcacggcga tctatgcggc ttctataaca acgtggaacg ggtgggcatg 9420

gattgcaggc gctgccctct acttgctctg cctgccggcg ctgcgtcgcg ggctttggag 9480

cggcgcaggg caacgagccg atcgctgatc gtggaaacga taggcctatg ccatgcgggt 9540

caaggcgact tccggcaagc tatacgcgcc ctaggagtgc ggttggaacg ttggcccagc 9600

cagatactcc cgatcacgag caggacgccg atgatttgaa gcgcactcag cgtctgatcc 9660

aagaacaacc atcctagcaa cacggcggtc cccgggctga gaaagcccag taaggaaaca 9720

actgtaggtt cgagtcgcga gatcccccgg aaccaaagga agtaggttaa acccgctccg 9780

atcaggccga gccacgccag gccgagaaca ttggttcctg taggcatcgg gattggcgga 9840

tcaaacacta aagctactgg aacgagcaga agtcctccgg ccgccagttg ccaggcggta 9900

aaggtgagca gaggcacggg aggttgccac ttgcgggtca gcacggttcc gaacgccatg 9960

gaaaccgccc ccgccaggcc cgctgcgacg ccgacaggat ctagcgctgc gtttggtgtc 10020

aacaccaaca gcgccacgcc cgcagttccg caaatagccc ccaggaccgc catcaatcgt 10080

atcgggctac ctagcagagc ggcagagatg aacacgacca tcagcggctg cacagcgcct 10140

accgtcgccg cgaccccgcc cggcaggcgg tagaccgaaa taaacaacaa gctccagaat 10200

agcgaaatat taagtgcgcc gaggatgaag atgcgcatcc accagattcc cgttggaatc 10260

tgtcggacga tcatcacgag caataaaccc gccggcaacg cccgcagcag cataccggcg 10320

acccctcggc ctcgctgttc gggctccacg aaaacgccgg acagatgcgc cttgtgagcg 10380

tccttggggc cgtcctcctg tttgaagacc gacagcccaa tgatctcgcc gtcgatgtag 10440

gcgccgaatg ccacggcatc tcgcaaccgt tcagcgaacg cctccatggg ctttttctcc 10500

tcgtgctcgt aaacggaccc gaacatctct ggagctttct tcagggccga caatcggatc 10560

tcgcggaaat cctgcacgtc ggccgctcca agccgtcgaa tctgagcctt aatcacaatt 10620

gtcaatttta atcctctgtt tatcggcagt tcgtagagcg cgccgtgcgt cccgagcgat 10680

actgagcgaa gcaagtgcgt cgagcagtgc ccgcttgttc ctgaaatgcc agtaaagcgc 10740

tggctgctga acccccagcc ggaactgacc ccacaaggcc ctagcgtttg caatgcacca 10800

ggtcatcatt gacccaggcg tgttccacca ggccgctgcc tcgcaactct tcgcaggctt 10860

cgccgacctg ctcgcgccac ttcttcacgc gggtggaatc cgatccgcac atgaggcgga 10920

aggtttccag cttgagcggg tacggctccc ggtgcgagct gaaatagtcg aacatccgtc 10980

gggccgtcgg cgacagcttg cggtacttct cccatatgaa tttcgtgtag tggtcgccag 11040

caaacagcac gacgatttcc tcgtcgatca ggacctggca acgggacgtt ttcttgccac 11100

ggtccaggac gcggaagcgg tgcagcagcg acaccgattc caggtgccca acgcggtcgg 11160

acgtgaagcc catcgccgtc gcctgtaggc gcgacaggca ttcctcggcc ttcgtgtaat 11220

accggccatt gatcgaccag cccaggtcct ggcaaagctc gtagaacgtg aaggtgatcg 11280

gctcgccgat aggggtgcgc ttcgcgtact ccaacacctg ctgccacacc agttcgtcat 11340

cgtcggcccg cagctcgacg ccggtgtagg tgatcttcac gtccttgttg acgtggaaaa 11400

tgaccttgtt ttgcagcgcc tcgcgcggga ttttcttgtt gcgcgtggtg aacagggcag 11460

agcgggccgt gtcgtttggc atcgctcgca tcgtgtccgg ccacggcgca atatcgaaca 11520

aggaaagctg catttccttg atctgctgct tcgtgtgttt cagcaacgcg gcctgcttgg 11580

-continued

cctcgctgac ctgttttgcc aggtcctcgc cggcggtttt tcgcttcttg gtcgtcatag 11640

ttcctcgcgt gtcgatggtc atcgacttcg ccaaacctgc cgcctcctgt tcgagacgac 11700

gcgaacgctc cacggcggcc gatggcgcgg gcagggcagg gggagccagt tgcacgctgt 11760

cgcgctcgat cttggccgta gcttgctgga ccatcgagcc gacggactgg aaggtttcgc 11820

ggggcgcacg catgacggtg cggcttgcga tggtttcggc atcctcggcg gaaaaccccg 11880

cgtcgatcag ttcttgcctg tatgccttcc ggtcaaacgt ccgattcatt caccctcctt 11940

gcgggattgc cccgactcac gccggggcaa tgtgccctta ttcctgattt gacccgcctg 12000

gtgccttggt gtccagataa tccaccttat cggcaatgaa gtcggtcccg tagaccgtct 12060

ggccgtcctt ctcgtacttg gtattccgaa tcttgccctg cacgaatacc agcgacccct 12120

tgcccaaata cttgccgtgg gcctcggcct gagagccaaa acacttgatg cggaagaagt 12180

cggtgcgctc ctgcttgtcg ccggtcgtgg ccgcgccaac ctttgcgatc cgcaagcgcg 12240

cggtcgccat cttcacgctg gaacagtacg tcgaggcggg catcatgacc cgcgagcaat 12300

acgaggtcat taaaagcgcc gtgattgatg atatagcggc ccggctgctc ctggttctcg 12360

cgcaccgaaa tgggtgactt caccccgcgc tctttgatcg tggcaccgat ttccgcgatg 12420

ctctccgggg aaaagccggg gttgtcggcc gtccgcggct gatgcggatc ttcgtcgatc 12480

aggtccaggt ccagctcgat agggccggaa ccgccctgag acgccgcagg agcgtccagg 12540

aggctcgaca ggtcgccgat gctatccaac cccaggccgg acggctgcgc cgcgcctgcg 12600

gcttcctgag cggccgcagc ggtgtttttc ttggtggtct tggcttgagc cgcagtcatt 12660

gggaaatctc catcttcgtg aacacgtaat cagccagggc gcgaacctct ttcgatgcct 12720

tgcgcgcggc cgttttcttg atcttccaga ccggcacacc ggatgcgagg gcatcggcga 12780

tgctgctgcg caggccaacg gtggccggaa tcatcatctt ggggtacgcg gccagcagct 12840

cggcttggtg gcgcgcgtgg cgcggattcc gcgcatcgac cttgctgggc accatgccaa 12900

ggaattgcag cttggcgttc ttctggcgca cgttcgcaat ggtcgtgacc atcttcttga 12960

tgccctggat gctgtacgcc tcaagctcga tgggggacag cacatagtcg gccgcgaaga 13020

gggcggccgc caggccgacg ccaagggtcg gggccgtgtc gatcaggcac acgtcgaagc 13080

cttggttcgc cagggccttg atgttcgccc cgaacagctc gcgggcgtcg tccagcgaca 13140

gccgttcggc gttcgccagt accgggttgg actcgatgag ggcgaggcgc gcggcctggc 13200

cgtcgccggc tgcgggtgcg gtttcggtcc agccgccggc agggacagcg ccgaacagct 13260

tgcttgcatg caggccggta gcaaagtcct tgagcgtgta ggacgcattg ccctgggggt 13320

ccaggtcgat cacggcaacc cgcaagccgc gctcgaaaaa gtcgaaggca agatgcacaa 13380

gggtcgaagt cttgccgacg ccgcctttct ggttggccgt gaccaaagtt ttcatcgttt 13440

ggtttcctgt tttttcttgg cgtccgcttc ccacttccgg acgatgtacg cctgatgttc 13500

cggcagaacc gccgttaccc gcgcgtaccc ctcgggcaag ttcttgtcct cgaacgcggc 13560

ccacacgcga tgcaccgctt gcgacactgc gcccctggtc agtcccagcg acgttgcgaa 13620

cgtcgcctgt ggcttcccat cgactaagac gccccgcgct atctcgatgg tctgctgccc 13680

cacttccagc ccctggatcg cctcctggaa ctggctttcg gtaagccgtt tcttcatgga 13740

taacacccat aatttgctcc gcgccttggt tgaacatagc ggtgacagcc gccagcacat 13800

gagagaagtt tagctaaaca tttctcgcac gtcaacacct ttagccgcta aaactcgtcc 13860

-continued

ttggcgtaac aaaacaaaag cccggaaacc gggctttcgt ctcttgccgc ttatggctct 13920

gcacccggct ccatcaccaa caggtcgcgc acgcgcttca ctcggttgcg gatcgacact 13980

gccagcccaa caaagccggt tgccgccgcc gccaggatcg cgccgatgat gccggccaca 14040

ccggccatcg cccaccaggt cgccgccttc cggttccatt cctgctggta ctgcttcgca 14100

atgctggacc tcggctcacc ataggctgac cgctcgatgg cgtatgccgc ttctcccctt 14160

ggcgtaaaac ccagcgccgc aggcggcatt gccatgctgc ccgccgcttt cccgaccacg 14220

acgcgcgcac caggcttgcg gtccagacct tcggccacgg cgagctgcgc aaggacataa 14280

tcagccgccg acttggctcc acgcgcctcg atcagctctt gcactcgcgc gaaatccttg 14340

gcctccacgg ccgccatgaa tcgcgcacgc ggcgaaggct ccgcagggcc 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggcg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccccc catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

-continued

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gcctggcccg caagaggggc tcgaagaccg acaacgaagc ggtgcgcgga ctgatccgcc 1740

agcaccgtcc gctcgtgctc atggggtcga tcaacgagaa gatttacctt gcggaactga 1800

aagccggtca cggcccggca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

cgggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggcttgc cctcgatgcc ggtgccgaga gggtcgtgag 2160

agagactgtc gaagccctga cgtgggccgg gcttcggcga gaggaaggga gagaaccaat 2220

gagcgatcat gccgtcaaca cgccggtcca tgccgccagg gcccacgggc accgagcacc 2280

tcgtgccgag ttctacgtct acttcgccgt cattctgctg ggcgccttcc cggtggcctt 2340

cgtgagctgg atcgtctcga cgatccgcca ccgcaggctt cccaagcgcg gccccttcgc 2400

gtccgcctgg ttcgatgcca aggcgatcac gccgctgatt ttccgcgcct gaccgcaggt 2460

caggttgcga cacgccattc gtcgtctccc caaggggcgg cggattaatc gggagggcat 2520

ggtgccttac cgtaacccac gccaccagta ggcaggagga acacgtgtcg tccggtggtc 2580

accaccacca ccaccaccac caccaccact aataggccgg ccctccgtcg cgggcggcac 2640

ccacgcccgc atcgattcca aggttcagcc attgagacgg ctccgcttcg cgcgcaagcg 2700

cgggttgggc cgactgcaag cggagaggga agcatggcac tgctcagctt cgagcgaaaa 2760

tatcgcgtgc cgggggggcac gctggtcggc ggaaacctgt tcgacttctg ggtcggccct 2820

ttctatgtcg gcttcttcgg ggttgcgacg tttttcttcg cggccctggg tatcattctg 2880

attgcctgga gtgccgtact ccagggtacc tggaaccccc aactcatctc tgtctacccg 2940

ccggcccttg aatatggcct gggaggtgca cccctcgcaa aaggcgggct gtggcagatc 3000

atcacgatct gcgccactgg tgccttcgtc agctgggcgc tgcgcgaagt cgaaatctgc 3060

cgtaagctgg gcatcgggta ccacatcccg ttcgccttcg cgttcgccat cctggcctac 3120

ctgacgctgg tgctgttccg cccggtgatg atgggcgcct ggggctatgc cttcccctac 3180

gggatctgga cgcacctcga ctgggtgtcg aacacgggct acacctacgg caacttccac 3240

tacaaccctg cccacatgat cgccatctcg ttcttcttca cgaacgcgct ggctctggcg 3300

ctgcacggcg cccttgtgct ctccgcggcc aaccccgaga agggcaagga aatgcggacg 3360

ccggatcacg aggatacgtt cttccgcgat ctggtcggct actcgatcgg gacgctcggc 3420

atccaccgcc tcggcctgct gctctcgctg agcgccgtct tcttcagcgc cctctgcatg 3480

atcattaccg gcaccatctg gttcgatcag tgggtcgact ggtggcaatg gtgggtgaag 3540

ctgccgtggt gggcgaacat cccgggaggc atcaatggct gagtatcaga acatcttctc 3600

ccaggtccag gtccgcggac cggccgacct ggggatgacc gaagacgtca acctggccaa 3660

-continued

ccgttcgggc gtcggtccct tctcgaccct gctcggctgg ttcggcaacg cccagctcgg 3720

cccgatctat ctcggctcgc tcggcgtcct gtccctcttc tcgggcctga tgtggttctt 3780

caccatcggg atctggttct ggtatcaggc gggctggaac ccggccgtct tcctgcgcga 3840

cctgttcttc ttctcgctcg agccgccggc acccgaatac ggtctgtcct tcgcggctcc 3900

gctgaaggaa ggcgggctgt ggctgatcgc gtcgttcttc atgttcgtcg cggtctggtc 3960

ctggtggggc cgcacctatc tccgcgctca ggcgctgggc atgggcaagc acaccgcctg 4020

ggcgttcctc tcggccatct ggctgtggat ggtgctgggc ttcatccgtc cgatcctcat 4080

ggggtcctgg tcggaagcgg ttccctacgg catcttctcg cacctcgact ggacgaacaa 4140

cttctcgctc gtccacggca acctgttcta caaccccttc cacggtctct cgatcgcctt 4200

cctctacggg tcggccctgc tcttcgcgat gcacggtgcg accatcctcg cggtctcccg 4260

cttcggcggc gagcgcgagc tggagcagat cgccgaccgc gggacggcag cggagcgggc 4320

cgccctcttc tggcgctgga ccatgggttt caacgccacg atggaaggca tccaccgctg 4380

ggccatctgg atggcggtcc tcgtgaccct caccggcggc atcgggatcc tgctctcggg 4440

cacggtcgtg gacaactggt acgtctgggg ccagaaccac ggcatggcgc cgctgaactg 4500

aggagcgatc acaatggctg acaagaccat cttcaacgat cacctcaaca ccaatccgaa 4560

gaccaacctt cgcctctggg tcgctttcca gatgatgaag ggtgcgggct gggctggcgg 4620

cgtgttcttc gggacgctcc ttctcatcgg gttcttccgg gtggtcgggc ggatgcttcc 4680

gatccaggag aaccaggctc cggcgccgaa catcaccggc gctctggaga ccgggatcga 4740

gctgatcaag catctcgtct gagacaagtc tcggggcagg gcggcgcgag gccgcccgct 4800

cctccaagtc cgggccatat cgccggcccg ggtccggggc gacaccacag cccggttccc 4860

ttcctgttgg cgacagggac ctggtgccgt gtggaagacc gcacggcacc cttttgacat 4920

tcacgggagg ctctgatgac caatcccacc ccgcgacccg aaaccccgct tttggatcgc 4980

gtctgctgcc cggccgacat gaaggcgctg agtgacgccg aactggagcg gctggccgac 5040

gaagtgcgtt ccgaggtcag tgataggggt agtttcttat tttaggcagt ttatatgaaa 5100

ttaagacatg cagatgtcac agtggatatt gaactggtct cgaaagctca atatccccca 5160

aagcacaagc acaaacttcg acatcatgca gaagcgtttc ccgaaccgcg tcttcgacgt 5220

gggcatcgcc gagcagcatg ccgtgacctt cgcggccggc ctcgccgggg ccgggatgaa 5280

gcccttctgc gcgatctatt cctcgttcct gcaacggggt tacgaccaga tcgcccatga 5340

cgtggcgctg cagaaccttc ccgtccgctt cgtgatcgac cgggcggggc tcgtgggggc 5400

cgatggcgcg acccatgcgg gggccttcga cgttggcttc atcacttcgc tgcccaacat 5460

gaccgtgatg gccgcggccg acgaggccga gctcatccac atgatcgcca ccgccgtggc 5520

cttcgacgag ggccccatcg ccttccgctt cccgcggggc gaggggggtgg gcgtcgagat 5580

gcccgagcgc gggacggtgc tggagcccgg ccggggccgc gtggtgcgcg aagggacgga 5640

tgtcgcgatc ctctccttcg gcgcgcatct gcacgaggcc ttgcaggcgg cgaaacttct 5700

cgaggccgag ggggtgagcg tgaccgtggc cgacgcccgc ttctcgcgcc cgctcgacac 5760

ggggctcatc gaccagctcg tgcgccatca cgcggcgctg gtaacggtgg agcaggggggc 5820

catgggcggc ttcggcgccc atgtcatgca ctatctcgcc aattccggcg gcttcgacgg 5880

gggcctcgcg ctccgggtca tgacgctgcc cgaccgcttc atcgagcagg cgagccccga 5940

-continued

ggacatgtat gccgatgcgg ggctgcgggc cgaggatatc aagcttggcg taatcatggt 6000

catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac atacgagccg 6060

gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca ttaattgcgt 6120

tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg 6180

gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ggtcttgcct tgctcgtcgg 6240

tgatgtactt caccagctcc gcgaagtcgc tcttcttgat ggagcgcatg gggacgtgct 6300

tggcaatcac gcgcaccccc cggccgtttt agcggctaaa aaagtcatgg ctctgccctc 6360

gggcggacca cgcccatcat gaccttgcca agctcgtcct gcttctcttc gatcttcgcc 6420

agcagggcga ggatcgtggc atcaccgaac cgcgccgtgc gcgggtcgtc ggtgagccag 6480

agtttcagca ggccgcccag gcggcccagg tcgccattga tgcgggccag ctcgcggacg 6540

tgctcatagt ccacgacgcc cgtgattttg tagccctggc cgacggccag caggtaggcc 6600

gacaggctca tgccggccgc cgccgccttt tcctcaatcg ctcttcgttc gtctggaagg 6660

cagtacacct tgataggtgg gctgcccttc ctggttggct tggtttcatc agccatccgc 6720

ttgccctcat ctgttacgcc ggcggtagcc ggccagcctc gcagagcagg attcccgttg 6780

agcaccgcca ggtgcgaata agggacagtg aagaaggaac acccgctcgc gggtgggcct 6840

acttcaccta tcctgcccgg ctgacgccgt tggatacacc aaggaaagtc tacacgaacc 6900

ctttggcaaa atcctgtata tcgtgcgaaa aaggatggat ataccgaaaa aatcgctata 6960

atgaccccga agcagggtta tgcagcggaa aagcgccacg cttcccgaag ggagaaaggc 7020

ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agcttccagg 7080

gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg 7140

atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt 7200

tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc 7260

tgattctgtg gataaccgta ttaccgcctt tgagtgagct gataccgctc gccgcagccg 7320

aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa gagcgccaga aggccgccag 7380

agaggccgag cgcggccgtg aggcttggac gctagggcag ggcatgaaaa agcccgtagc 7440

gggctgctac gggcgtctga cgcggtggaa agggggaggg gatgttgtct acatggctct 7500

gctgtagtga gtgggttgcg ctccggcagc ggtcctgatc aatcgtcacc ctttctcggt 7560

ccttcaacgt tcctgacaac gagcctcctt ttcgccaatc catcgacaat caccgcgagt 7620

ccctgctcga acgctgcgtc cggaccggct tcgtcgaagg cgtctatcgc ggcccgcaac 7680

agcggcgaga gcggagcctg ttcaacggtg ccgccgcgct cgccggcatc gctgtcgccg 7740

gcctgctcct caagcacggc cccaacagtg aagtagctga ttgtcatcag cgcattgacg 7800

gcgtccccgg ccgaaaaacc cgcctcgcag aggaagcgaa gctgcgcgtc ggccgtttcc 7860

atctgcggtg cgcccggtcg cgtgccggca tggatgcgcg cgccatcgcg gtaggcgagc 7920

agcgcctgcc tgaagctgcg ggcattcccg atcagaaatg agcgccagtc gtcgtcggct 7980

ctcggcaccg aatgcgtatg attctccgcc agcatggctt cggccagtgc gtcgagcagc 8040

gcccgcttgt tcctgaagtg ccagtaaagc gccggctgct gaacccccaa ccgttccgcc 8100

agtttgcgtg tcgtcagacc gtctacgccg acctcgttca acaggtccag ggcggcacgg 8160

atcactgtat tcggctgcaa ctttgtcatg cttgacactt tatcactgat aaacataata 8220

-continued

tgtccaccaa cttatcagtg ataaagaatc cgcgcgttca atcggaccag cggaggctgg 8280

tccggaggcc agacatgaaa cccaacatac ccctgatcgt aattctgagc actgtcgcgc 8340

tcgacgctgt cggcatcggc ctgattatgc cggtgctgcc gggcctcctg cgcgatctgg 8400

ttcactcgaa cgacgtcacc gcccactatg gcattctgct ggcgctgtat gcgttggtgc 8460

aatttgcctg cgcacctgtg ctgggcgcgc tgtcggatcg tttcgggcgg cggccaatct 8520

tgctcgtctc gctggccggc gccactgtcg actacgccat catggcgaca gcgcctttcc 8580

tttgggttct ctatatcggg cggatcgtgg ccggcatcac cggggcgact ggggcggtag 8640

ccggcgctta tattgccgat atcactgatg gcgatgagcg cgcgcggcac ttcggcttca 8700

tgagcgcctg tttcgggttc gggatggtcg cgggacctgt gctcggtggg ctgatgggcg 8760

gtttctcccc ccacgctccg ttcttcgccg cggcagcctt gaacggcctc aatttcctga 8820

cgggctgttt ccttttgccg gagtcgcaca aaggcgaacg ccggccgtta cgccgggagg 8880

ctctcaaccc gctcgcttcg ttccggtggg cccggggcat gaccgtcgtc gccgccctga 8940

tggcggtctt cttcatcatg caacttgtcg gacaggtgcc ggccgcgctt tgggtcattt 9000

tcggcgagga tcgctttcac tgggacgcga ccacgatcgg catttcgctt gccgcatttg 9060

gcattctgca ttcactcgcc caggcaatga tcaccggccc tgtagccgcc cggctcggcg 9120

aaaggcgggc actcatgctc ggaatgattg ccgacggcac aggctacatc ctgcttgcct 9180

tcgcgacacg gggatggatg gcgttcccga tcatggtcct gcttgcttcg ggtggcatcg 9240

gaatgccggc gctgcaagca atgttgtcca ggcaggtgga tgaggaacgt caggggcagc 9300

tgcaaggctc actggcggcg ctcaccagcc tgacctcgat cgtcggaccc ctcctcttca 9360

cggcgatcta tgcggcttct ataacaacgt ggaacgggtg ggcatggatt gcaggcgctg 9420

ccctctactt gctctgcctg ccggcgctgc gtcgcgggct ttggagcggc gcagggcaac 9480

gagccgatcg ctgatcgtgg aaacgatagg cctatgccat gcgggtcaag gcgacttccg 9540

gcaagctata cgcgccctag gagtgcggtt ggaacgttgg cccagccaga tactcccgat 9600

cacgagcagg acgccgatga tttgaagcgc actcagcgtc tgatccaaga acaaccatcc 9660

tagcaacacg gcggtccccg ggctgagaaa gcccagtaag gaaacaactg taggttcgag 9720

tcgcgagatc ccccggaacc aaaggaagta ggttaaaccc gctccgatca ggccgagcca 9780

cgccaggccg agaacattgg ttcctgtagg catcgggatt ggcggatcaa acactaaagc 9840

tactggaacg agcagaagtc ctccggccgc cagttgccag gcggtaaagg tgagcagagg 9900

cacgggaggt tgccacttgc gggtcagcac ggttccgaac gccatggaaa ccgcccccgc 9960

caggcccgct gcgacgccga caggatctag cgctgcgttt ggtgtcaaca ccaacagcgc 10020

cacgcccgca gttccgcaaa tagcccccag gaccgccatc aatcgtatcg ggctacctag 10080

cagagcggca gagatgaaca cgaccatcag cggctgcaca gcgcctaccg tcgccgcgac 10140

cccgcccggc aggcggtaga ccgaaataaa caacaagctc cagaatagcg aaatattaag 10200

tgcgccgagg atgaagatgc gcatccacca gattcccgtt ggaatctgtc ggacgatcat 10260

cacgagcaat aaacccgccg gcaacgcccg cagcagcata ccggcgaccc ctcggcctcg 10320

ctgttcgggc tccacgaaaa cgccggacag atgcgccttg tgagcgtcct tggggccgtc 10380

ctcctgtttg aagaccgaca gcccaatgat ctcgccgtcg atgtaggcgc cgaatgccac 10440

ggcatctcgc aaccgttcag cgaacgcctc catgggcttt ttctcctcgt gctcgtaaac 10500

-continued

ggacccgaac atctctggag ctttcttcag ggccgacaat cggatctcgc ggaaatcctg 10560

cacgtcggcc gctccaagcc gtcgaatctg agccttaatc acaattgtca attttaatcc 10620

tctgtttatc ggcagttcgt agagcgcgcc gtgcgtcccg agcgatactg agcgaagcaa 10680

gtgcgtcgag cagtgcccgc ttgttcctga aatgccagta aagcgctggc tgctgaaccc 10740

ccagccggaa ctgaccccac aaggccctag cgtttgcaat gcaccaggtc atcattgacc 10800

caggcgtgtt ccaccaggcc gctgcctcgc aactcttcgc aggcttcgcc gacctgctcg 10860

cgccacttct tcacgcgggt ggaatccgat ccgcacatga ggcggaaggt ttccagcttg 10920

agcgggtacg gctcccggtg cgagctgaaa tagtcgaaca tccgtcgggc cgtcggcgac 10980

agcttgcggt acttctccca tatgaatttc gtgtagtggt cgccagcaaa cagcacgacg 11040

atttcctcgt cgatcaggac ctggcaacgg gacgttttct tgccacggtc caggacgcgg 11100

aagcggtgca gcagcgacac cgattccagg tgcccaacgc ggtcggacgt gaagcccatc 11160

gccgtcgcct gtaggcgcga caggcattcc tcggccttcg tgtaataccg gccattgatc 11220

gaccagccca ggtcctggca aagctcgtag aacgtgaagg tgatcggctc gccgataggg 11280

gtgcgcttcg cgtactccaa cacctgctgc cacaccagtt cgtcatcgtc ggcccgcagc 11340

tcgacgccgg tgtaggtgat cttcacgtcc ttgttgacgt ggaaaatgac cttgttttgc 11400

agcgcctcgc gcgggatttt cttgttgcgc gtggtgaaca gggcagagcg ggccgtgtcg 11460

tttggcatcg ctcgcatcgt gtccggccac ggcgcaatat cgaacaagga aagctgcatt 11520

tccttgatct gctgcttcgt gtgtttcagc aacgcggcct gcttggcctc gctgacctgt 11580

tttgccaggt cctcgccggc ggtttttcgc ttcttggtcg tcatagttcc tcgcgtgtcg 11640

atggtcatcg acttcgccaa acctgccgcc tcctgttcga gacgacgcga acgctccacg 11700

gcggccgatg gcgcgggcag ggcaggggga gccagttgca cgctgtcgcg ctcgatcttg 11760

gccgtagctt gctggaccat cgagccgacg gactggaagg tttcgcgggg cgcacgcatg 11820

acggtgcggc ttgcgatggt ttcggcatcc tcggcggaaa accccgcgtc gatcagttct 11880

tgcctgtatg ccttccggtc aaacgtccga ttcattcacc ctccttgcgg gattgccccg 11940

actcacgccg gggcaatgtg cccttattcc tgatttgacc cgcctggtgc cttggtgtcc 12000

agataatcca ccttatcggc aatgaagtcg gtcccgtaga ccgtctggcc gtccttctcg 12060

tacttggtat tccgaatctt gccctgcacg aataccagcg acccccttgcc caaatacttg 12120

ccgtgggcct cggcctgaga gccaaaaacac ttgatgcgga agaagtcggt gcgctcctgc 12180

ttgtcgccgg tcgtggccgc gccaaccttt gcgatccgca agcgcgcggt cgccatcttc 12240

acgctggaac agtacgtcga ggcgggcatc atgacccgcg agcaatacga ggtcattaaa 12300

agcgccgtga ttgatgatat agcggcccgg ctgctcctgg ttctcgcgca ccgaaatggg 12360

tgacttcacc ccgcgctctt tgatcgtggc accgatttcc gcgatgctct ccggggaaaa 12420

gccggggttg tcggccgtcc gcggctgatg cggatcttcg tcgatcaggt ccaggtccag 12480

ctcgataggg ccggaaccgc cctgagacgc cgcaggagcg tccaggaggc tcgacaggtc 12540

gccgatgcta tccaacccca ggccggacgg ctgcgccgcg cctgcggctt cctgagcggc 12600

cgcagcggtg tttttcttgg tggtcttggc ttgagccgca gtcattggga aatctccatc 12660

ttcgtgaaca cgtaatcagc cagggcgcga acctctttcg atgccttgcg cgcggccgtt 12720

ttcttgatct tccagaccgg cacaccggat gcgagggcat cggcgatgct gctgcgcagg 12780

-continued

ccaacggtgg ccggaatcat catcttgggg tacgcggcca gcagctcggc ttggtggcgc 12840

gcgtggcgcg gattccgcgc atcgaccttg ctgggcacca tgccaaggaa ttgcagcttg 12900

gcgttcttct ggcgcacgtt cgcaatggtc gtgaccatct tcttgatgcc ctggatgctg 12960

tacgcctcaa gctcgatggg ggacagcaca tagtcggccg cgaagagggc ggccgccagg 13020

ccgacgccaa gggtcggggc cgtgtcgatc aggcacacgt cgaagccttg gttcgccagg 13080

gccttgatgt tcgccccgaa cagctcgcgg gcgtcgtcca gcgacagccg ttcggcgttc 13140

gccagtaccg ggttggactc gatgagggcg aggcgcgcgg cctggccgtc gccggctgcg 13200

ggtgcggttt cggtccagcc gccggcaggg acagcgccga acagcttgct tgcatgcagg 13260

ccggtagcaa agtccttgag cgtgtaggac gcattgccct gggggtccag gtcgatcacg 13320

gcaacccgca agccgcgctc gaaaaagtcg aaggcaagat gcacaagggt cgaagtcttg 13380

ccgacgccgc ctttctggtt ggccgtgacc aaagttttca tcgtttggtt tcctgttttt 13440

tcttggcgtc cgcttcccac ttccggacga tgtacgcctg atgttccggc agaaccgccg 13500

ttacccgcgc gtacccctcg ggcaagttct tgtcctcgaa cgcggcccac acgcgatgca 13560

ccgcttgcga cactgcgccc ctggtcagtc ccagcgacgt tgcgaacgtc gcctgtggct 13620

tcccatcgac taagacgccc cgcgctatct cgatggtctg ctgccccact tccagcccct 13680

ggatcgcctc ctggaactgg ctttcggtaa gccgtttctt catggataac acccataatt 13740

tgctccgcgc cttggttgaa catagcggtg acagccgcca gcacatgaga gaagtttagc 13800

taaacatttc tcgcacgtca acacctttag ccgctaaaac tcgtccttgg cgtaacaaaa 13860

caaaagcccg gaaaccgggc tttcgtctct tgccgcttat ggctctgcac ccggctccat 13920

caccaacagg tcgcgcacgc gcttcactcg gttgcggatc gacactgcca gcccaacaaa 13980

gccggttgcc gccgccgcca ggatcgcgcc gatgatgccg gccacaccgg ccatcgccca 14040

ccaggtcgcc gccttccggt tccattcctg ctggtactgc ttcgcaatgc tggacctcgg 14100

ctcaccatag gctgaccgct cgatggcgta tgccgcttct ccccttggcg taaaacccag 14160

cgccgcaggc ggcattgcca tgctgcccgc cgctttcccg accacgacgc gcgcaccagg 14220

cttgcggtcc agaccttcgg ccacggcgag ctgcgcaagg acataatcag ccgccgactt 14280

ggctccacgc gcctcgatca gctcttgcac tcgcgcgaaa tccttggcct ccacggccgc 14340

catgaatcgc gcacgcggcg aaggctccgc 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

-continued

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccct catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gcctggcccg caagaggggc tcgaagaccg acaacgaagc ggtgcgcgga ctgatccgcc 1740

agcaccgtcc gctcgtgctc atggggtcga tcaacgagaa gatttacctt gcggaactga 1800

aagccggtca cggcccggca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

cgggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggcttgc cctcgatgcc ggtgccgaga gggtcgtgag 2160

agagactgtc gaagccctga cgtgggccgg gcttcggcga gaggaaggga gagaaccaat 2220

gagcgatcat gccgtcaaca cgccggtcca tgccgccagg gcccacgggc accgagcacc 2280

tcgtgccgag ttctacgtct acttcgccgt cattctgctg ggcgccttcc cggtggcctt 2340

cgtgagctgg atcgtctcga cgatccgcca ccgcaggctt cccaagcgcg gccccttcgc 2400

gtccgcctgg ttcgatgcca aggcgatcac gccgctgatt ttccgcgcct gaccgcaggt 2460

caggttgcga cacgccattc gtcgtctccc caaggggcgg cggattaatc gggagggcat 2520

ggtgccttac cgtaacccac gccaccagta ggcaggagga acacgtgtcg tccggtggtc 2580

accaccacca ccaccaccac caccaccacc accaccacta ataggccggc cctccgtcgc 2640

-continued

```
gggcggcacc cacgcccgca tcgattccaa ggttcagcca ttgagacggc tccgcttcgc   2700

gcgcaagcgc gggttgggcc gactgcaagc ggagagggaa gcatggcact gctcagcttc   2760

gagcgaaaat atcgcgtgcc gggggcacg ctggtcggcg gaaacctgtt cgacttctgg   2820

gtcggccctt tctatgtcgg cttcttcggg gttgcgacgt ttttcttcgc ggccctgggt   2880

atcattctga ttgcctggag tgccgtactc cagggtacct ggaaccccca actcatctct   2940

gtctacccgc cggcccttga atatggcctg ggaggtgcac ccctcgcaaa aggcgggctg   3000

tggcagatca tcacgatctg cgccactggt gccttcgtca gctgggcgct gcgcgaagtc   3060

gaaatctgcc gtaagctggg catcgggtac cacatcccgt tcgccttcgc gttcgccatc   3120

ctggcctacc tgacgctggt gctgttccgc ccggtgatga tgggcgcctg gggctatgcc   3180

ttcccctacg ggatctggac gcacctcgac tgggtgtcga acacgggcta cacctacggc   3240

aacttccact acaaccctgc ccacatgatc gccatctcgt tcttcttcac gaacgcgctg   3300

gctctggcgc tgcacggcgc ccttgtgctc tccgcggcca accccgagaa gggcaaggaa   3360

atgcggacgc cggatcacga ggatacgttc ttccgcgatc tggtcggcta ctcgatcggg   3420

acgctcggca tccaccgcct cggcctgctg ctctcgctga gcgccgtctt cttcagcgcc   3480

ctctgcatga tcattaccgg caccatctgg ttcgatcagt gggtcgactg gtggcaatgg   3540

tgggtgaagc tgccgtggtg ggcgaacatc ccgggaggca tcaatggctg agtatcagaa   3600

catcttctcc caggtccagg tccgcggacc ggccgacctg gggatgaccg aagacgtcaa   3660

cctggccaac cgttcgggcg tcggtccctt ctcgaccctg ctcggctggt tcggcaacgc   3720

ccagctcggc ccgatctatc tcggctcgct cggcgtcctg tccctcttct cgggcctgat   3780

gtggttcttc accatcggga tctggttctg gtatcaggcg ggctggaacc cggccgtctt   3840

cctgcgcgac ctgttcttct tctcgctcga gccgccggca cccgaatacg gtctgtcctt   3900

cgcggctccg ctgaaggaag gcgggctgtg gctgatcgcg tcgttcttca tgttcgtcgc   3960

ggtctggtcc tggtggggcc gcacctatct ccgcgctcag gcgctgggca tgggcaagca   4020

caccgcctgg gcgttcctct cggccatctg gctgtggatg gtgctgggct tcatccgtcc   4080

gatcctcatg gggtcctggt cggaagcggt tccctacggc atcttctcgc acctcgactg   4140

gacgaacaac ttctcgctcg tccacggcaa cctgttctac aacccccttcc acggtctctc   4200

gatcgccttc ctctacgggt cggccctgct cttcgcgatg cacggtgcga ccatcctcgc   4260

ggtctcccgc ttcggcggcg agcgcgagct ggagcagatc gccgaccgcg ggacggcagc   4320

ggagcgggcc gccctcttct ggcgctggac catgggtttc aacgccacga tggaaggcat   4380

ccaccgctgg gccatctgga tggcggtcct cgtgaccctc accggcggca tcgggatcct   4440

gctctcgggc acggtcgtgg acaactggta cgtctggggc cagaaccacg gcatggcgcc   4500

gctgaactga ggagcgatca caatggctga caagaccatc ttcaacgatc acctcaacac   4560

caatccgaag accaaccttc gcctctgggt cgctttccag atgatgaagg gtgcgggctg   4620

ggctggcggc gtgttcttcg ggacgctcct tctcatcggg ttcttccggg tggtcgggcg   4680

gatgcttccg atccaggaga accaggctcc ggcgccgaac atcaccggcg ctctggagac   4740

cgggatcgag ctgatcaagc atctcgtctg agacaagtct cggggcaggg cggcgcgagg   4800

ccgcccgctc ctccaagtcc gggccatatc gccggcccgg gtccggggcg acaccacagc   4860

ccggttccct tcctgttggc gacagggacc tggtgccgtg tggaagaccg cacggcaccc   4920
```

-continued

ttttgacatt cacgggaggc tctgatgacc aatcccaccc cgcgacccga aaccccgctt 4980

ttggatcgcg tctgctgccc ggccgacatg aaggcgctga gtgacgccga actggagcgg 5040

ctggccgacg aagtgcgttc cgaggtcagt gataggggta gtttcttatt ttaggcagtt 5100

tatatgaaat taagacatgc agatgtcaca gtggatattg aactggtctc gaaagctcaa 5160

tatcccccaa agcacaagca caaacttcga catcatgcag aagcgtttcc cgaaccgcgt 5220

cttcgacgtg ggcatcgccg agcagcatgc cgtgaccttc gcggccggcc tcgccggggc 5280

cgggatgaag cccttctgcg cgatctattc ctcgttcctg caacggggtt acgaccagat 5340

cgcccatgac gtggcgctgc agaaccttcc cgtccgcttc gtgatcgacc gggcggggct 5400

cgtgggggcc gatggcgcga cccatgcggg ggccttcgac gttggcttca tcacttcgct 5460

gcccaacatg accgtgatgg ccgcggccga cgaggccgag ctcatccaca tgatcgccac 5520

cgccgtggcc ttcgacgagg gccccatcgc cttccgcttc ccgcggggcg agggggtggg 5580

cgtcgagatg cccgagcgcg ggacggtgct ggagcccggc cggggccgcg tggtgcgcga 5640

agggacggat gtcgcgatcc tctccttcgg cgcgcatctg cacgaggcct tgcaggcggc 5700

gaaacttctc gaggccgagg gggtgagcgt gaccgtggcc gacgcccgct tctcgcgccc 5760

gctcgacacg gggctcatcg accagctcgt gcgccatcac gcggcgctgg taacggtgga 5820

gcaggggcc atgggcggct tcggcgccca tgtcatgcac tatctcgcca attccggcgg 5880

cttcgacggg ggcctcgcgc tccgggtcat gacgctgccc gaccgcttca tcgagcaggc 5940

gagccccgag gacatgtatg ccgatgcggg gctgcgggcc gaggatatca agcttggcgt 6000

aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt ccacacaaca 6060

tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat 6120

taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt 6180

aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctcg gtcttgcctt 6240

gctcgtcggt gatgtacttc accagctccg cgaagtcgct cttcttgatg gagcgcatgg 6300

ggacgtgctt ggcaatcacg cgcacccccc ggccgtttta gcggctaaaa aagtcatggc 6360

tctgccctcg ggcggaccac gcccatcatg accttgccaa gctcgtcctg cttctcttcg 6420

atcttcgcca gcagggcgag gatcgtggca tcaccgaacc gcgccgtgcg cgggtcgtcg 6480

gtgagccaga gtttcagcag gccgcccagg cggcccaggt cgccattgat gcgggccagc 6540

tcgcggacgt gctcatagtc cacgacgccc gtgattttgt agccctggcc gacggccagc 6600

aggtaggccg acaggctcat gccggccgcc gccgcctttt cctcaatcgc tcttcgttcg 6660

tctggaaggc agtacacctt gataggtggg ctgcccttcc tggttggctt ggtttcatca 6720

gccatccgct tgccctcatc tgttacgccg gcggtagccg gccagcctcg cagagcagga 6780

ttcccgttga gcaccgccag gtgcgaataa gggacagtga agaaggaaca cccgctcgcg 6840

ggtgggccta cttcacctat cctgcccggc tgacgccgtt ggatacacca aggaaagtct 6900

acacgaaccc tttggcaaaa tcctgtatat cgtgcgaaaa aggatggata taccgaaaaa 6960

atcgctataa tgaccccgaa gcagggttat gcagcggaaa agcgccacgc ttcccgaagg 7020

gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga 7080

gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 7140

tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa 7200

-continued

```
cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc   7260

gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg   7320

ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgccagaa   7380

ggccgccaga gaggccgagc gcggccgtga ggcttggacg ctagggcagg gcatgaaaaa   7440

gcccgtagcg ggctgctacg ggcgtctgac gcggtggaaa gggggagggg atgttgtcta   7500

catggctctg ctgtagtgag tgggttgcgc tccggcagcg gtcctgatca atcgtcaccc   7560

tttctcggtc cttcaacgtt cctgacaacg agcctccttt tcgccaatcc atcgacaatc   7620

accgcgagtc cctgctcgaa cgctgcgtcc ggaccggctt cgtcgaaggc gtctatcgcg   7680

gcccgcaaca gcggcgagag cggagcctgt tcaacggtgc cgccgcgctc gccggcatcg   7740

ctgtcgccgg cctgctcctc aagcacggcc ccaacagtga agtagctgat tgtcatcagc   7800

gcattgacgg cgtccccggc cgaaaaaccc gcctcgcaga ggaagcgaag ctgcgcgtcg   7860

gccgtttcca tctgcggtgc gcccggtcgc gtgccggcat ggatgcgcgc gccatcgcgg   7920

taggcgagca gcgcctgcct gaagctgcgg gcattcccga tcagaaatga gcgccagtcg   7980

tcgtcggctc tcggcaccga atgcgtatga ttctccgcca gcatggcttc ggccagtgcg   8040

tcgagcagcg cccgcttgtt cctgaagtgc cagtaaagcg ccggctgctg aacccccaac   8100

cgttccgcca gtttgcgtgt cgtcagaccg tctacgccga cctcgttcaa caggtccagg   8160

gcggcacgga tcactgtatt cggctgcaac tttgtcatgc ttgacacttt atcactgata   8220

aacataatat gtccaccaac ttatcagtga taaagaatcc gcgcgttcaa tcggaccagc   8280

ggaggctggt ccggaggcca gacatgaaac ccaacatacc cctgatcgta attctgagca   8340

ctgtcgcgct cgacgctgtc ggcatcggcc tgattatgcc ggtgctgccg ggcctcctgc   8400

gcgatctggt tcactcgaac 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

gggcggtagc cggcgcttat attgccgata tcactgatgg cgatgagcgc gcgcggcact   8700

tcggcttcat gagcgcctgt ttcgggttcg ggatggtcgc gggacctgtg ctcggtgggc   8760

tgatgggcgg tttctccccc cacgctccgt tcttcgccgc ggcagccttg aacggcctca   8820

atttcctgac gggctgtttc cttttgccgg agtcgcacaa aggcgaacgc cggccgttac   8880

gccgggaggc tctcaacccg ctcgcttcgt tccggtgggc ccggggcatg accgtcgtcg   8940

ccgccctgat ggcggtcttc ttcatcatgc aacttgtcgg acaggtgccg gccgcgcttt   9000

gggtcatttt cggcgaggat cgctttcact gggacgcgac cacgatcggc atttcgcttg   9060

ccgcatttgg cattctgcat tcactcgccc aggcaatgat caccggccct gtagccgccc   9120

ggctcggcga aaggcgggca ctcatgctcg gaatgattgc cgacggcaca ggctacatcc   9180

tgcttgcctt cgcgacacgg ggatggatgg cgttcccgat catggtcctg cttgcttcgg   9240

gtggcatcgg aatgccggcg ctgcaagcaa tgttgtccag gcaggtggat gaggaacgtc   9300

aggggcagct gcaaggctca ctggcggcgc tcaccagcct gacctcgatc gtcggacccc   9360

tcctcttcac ggcgatctat gcggcttcta taacaacgtg gaacgggtgg gcatggattg   9420

caggcgctgc cctctacttg ctctgcctgc cggcgctgcg tcgcgggctt tggagcggcg   9480
```

-continued

cagggcaacg agccgatcgc tgatcgtgga aacgataggc ctatgccatg cgggtcaagg 9540

cgacttccgg caagctatac gcgccctagg agtgcggttg gaacgttggc ccagccagat 9600

actcccgatc acgagcagga cgccgatgat ttgaagcgca ctcagcgtct gatccaagaa 9660

caaccatcct agcaacacgg cggtccccgg 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 aggcccgctg cgacgccgac aggatctagc gctgcgtttg gtgtcaacac 10020

caacagcgcc acgcccgcag ttccgcaaat agcccccagg accgccatca atcgtatcgg 10080

gctacctagc agagcggcag agatgaacac gaccatcagc ggctgcacag cgcctaccgt 10140

cgccgcgacc ccgcccggca ggcggtagac cgaaataaac aacaagctcc agaatagcga 10200

aatattaagt gcgccgagga tgaagatgcg catccaccag attcccgttg gaatctgtcg 10260

gacgatcatc acgagcaata aacccgccgg caacgcccgc agcagcatac cggcgacccc 10320

tcggcctcgc tgttcgggct ccacgaaaac gccggacaga tgcgccttgt gagcgtcctt 10380

ggggccgtcc tcctgtttga agaccgacag cccaatgatc tcgccgtcga tgtaggcgcc 10440

gaatgccacg gcatctcgca accgttcagc gaacgcctcc atgggctttt tctcctcgtg 10500

ctcgtaaacg gacccgaaca tctctggagc tttcttcagg gccgacaatc ggatctcgcg 10560

gaaatcctgc acgtcggccg ctccaagccg tcgaatctga gccttaatca caattgtcaa 10620

ttttaatcct ctgtttatcg gcagttcgta gagcgcgccg tgcgtcccga gcgatactga 10680

gcgaagcaag tgcgtcgagc agtgcccgct tgttcctgaa atgccagtaa agcgctggct 10740

gctgaacccc cagccggaac tgaccccaca aggccctagc gtttgcaatg caccaggtca 10800

tcattgaccc aggcgtgttc caccaggccg ctgcctcgca actcttcgca ggcttcgccg 10860

acctgctcgc gccacttctt cacgcgggtg gaatccgatc cgcacatgag gcggaaggtt 10920

tccagcttga gcgggtacgg ctcccggtgc gagctgaaat agtcgaacat ccgtcgggcc 10980

gtcggcgaca gcttgcggta cttctcccat atgaatttcg tgtagtggtc gccagcaaac 11040

agcacgacga tttcctcgtc gatcaggacc tggcaacggg acgttttctt gccacggtcc 11100

aggacgcgga agcggtgcag cagcgacacc gattccaggt gcccaacgcg gtcggacgtg 11160

aagcccatcg ccgtcgcctg taggcgcgac aggcattcct cggccttcgt gtaataccgg 11220

ccattgatcg accagcccag gtcctggcaa agctcgtaga acgtgaaggt gatcggctcg 11280

ccgatagggg tgcgcttcgc gtactccaac acctgctgcc acaccagttc gtcatcgtcg 11340

gcccgcagct cgacgccggt gtaggtgatc ttcacgtcct tgttgacgtg gaaaatgacc 11400

ttgttttgca gcgcctcgcg cgggattttc ttgttgcgcg tggtgaacag ggcagagcgg 11460

gccgtgtcgt ttggcatcgc tcgcatcgtg tccggccacg gcgcaatatc gaacaaggaa 11520

agctgcattt ccttgatctg ctgcttcgtg tgtttcagca acgcggcctg cttggcctcg 11580

ctgacctgtt ttgccaggtc ctcgccggcg gtttttcgct tcttggtcgt catagttcct 11640

cgcgtgtcga tggtcatcga cttcgccaaa cctgccgcct cctgttcgag acgacgcgaa 11700

cgctccacgg cggccgatgg cgcgggcagg gcaggggggag ccagttgcac gctgtcgcgc 11760

-continued

tcgatcttgg ccgtagcttg ctggaccatc gagccgacgg actggaaggt ttcgcggggc 11820

gcacgcatga cggtgcggct tgcgatggtt tcggcatcct cggcggaaaa ccccgcgtcg 11880

atcagttctt gcctgtatgc cttccggtca aacgtccgat tcattcaccc tccttgcggg 11940

attgccccga ctcacgccgg ggcaatgtgc ccttattcct gatttgaccc gcctggtgcc 12000

ttggtgtcca gataatccac cttatcggca atgaagtcgg tcccgtagac cgtctggccg 12060

tccttctcgt acttggtatt ccgaatcttg ccctgcacga ataccagcga ccccttgccc 12120

aaatacttgc cgtgggcctc ggcctgagag ccaaaacact tgatgcggaa gaagtcggtg 12180

cgctcctgct tgtcgccggt cgtggccgcg ccaacctttg cgatccgcaa gcgcgcggtc 12240

gccatcttca cgctggaaca gtacgtcgag gcgggcatca tgacccgcga gcaatacgag 12300

gtcattaaaa gcgccgtgat tgatgatata gcggcccggc tgctcctggt tctcgcgcac 12360

cgaaatgggt gacttcaccc cgcgctcttt gatcgtggca ccgatttccg cgatgctctc 12420

cggggaaaag ccggggttgt cggccgtccg cggctgatgc ggatcttcgt cgatcaggtc 12480

caggtccagc tcgatagggc cggaaccgcc ctgagacgcc gcaggagcgt ccaggaggct 12540

cgacaggtcg ccgatgctat ccaaccccag gccggacggc tgcgccgcgc ctgcggcttc 12600

ctgagcggcc gcagcggtgt ttttcttggt ggtcttggct tgagccgcag tcattgggaa 12660

atctccatct tcgtgaacac gtaatcagcc agggcgcgaa cctctttcga tgccttgcgc 12720

gcggccgttt tcttgatctt ccagaccggc acaccggatg cgagggcatc ggcgatgctg 12780

ctgcgcaggc caacggtggc cggaatcatc atcttggggt acgcggccag cagctcggct 12840

tggtggcgcg cgtggcgcgg attccgcgca tcgaccttgc tgggcaccat gccaaggaat 12900

tgcagcttgg cgttcttctg gcgcacgttc gcaatggtcg tgaccatctt cttgatgccc 12960

tggatgctgt acgcctcaag ctcgatgggg gacagcacat agtcggccgc gaagagggcg 13020

gccgccaggc cgacgccaag ggtcggggcc gtgtcgatca ggcacacgtc gaagccttgg 13080

ttcgccaggg ccttgatgtt cgccccgaac agctcgcggg cgtcgtccag cgacagccgt 13140

tcggcgttcg ccagtaccgg gttggactcg atgagggcga ggcgcgcggc ctggccgtcg 13200

ccggctgcgg gtgcggtttc ggtccagccg ccggcaggga cagcgccgaa cagcttgctt 13260

gcatgcaggc cggtagcaaa gtccttgagc gtgtaggacg cattgccctg ggggtccagg 13320

tcgatcacgg caacccgcaa gccgcgctcg aaaaagtcga aggcaagatg cacaagggtc 13380

gaagtcttgc cgacgccgcc tttctggttg gccgtgacca aagttttcat cgtttggttt 13440

cctgtttttt cttggcgtcc gcttcccact tccggacgat gtacgcctga tgttccggca 13500

gaaccgccgt tacccgcgcg taccctcgg gcaagttctt gtcctcgaac gcggcccaca 13560

cgcgatgcac cgcttgcgac actgcgcccc tggtcagtcc cagcgacgtt gcgaacgtcg 13620

cctgtggctt cccatcgact aagacgcccc gcgctatctc gatggtctgc tgccccactt 13680

ccagcccctg gatcgcctcc tggaactggc tttcggtaag ccgtttcttc atggataaca 13740

cccataattt gctccgcgcc ttggttgaac atagcggtga cagccgccag cacatgagag 13800

aagtttagct aaacatttct cgcacgtcaa caccttagc cgctaaaact cgtccttggc 13860

gtaacaaaac aaaagcccgg aaaccgggct ttcgtctctt gccgcttatg gctctgcacc 13920

cggctccatc accaacaggt cgcgcacgcg cttcactcgg ttgcggatcg acactgccag 13980

cccaacaaag ccggttgccg ccgccgccag gatcgcgccg atgatgccgg ccacaccggc 14040

-continued

catcgcccac caggtcgccg ccttccggtt ccattcctgc tggtactgct tcgcaatgct 14100

ggacctcggc tcaccatagg ctgaccgctc gatggcgtat gccgcttctc cccttggcgt 14160

aaaacccagc gccgcaggcg gcattgccat gctgcccgcc gctttcccga ccacgacgcg 14220

cgcaccaggc ttgcggtcca gaccttcggc 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggcg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccccc catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

-continued

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gcctggcccg caagaggggc tcgaagaccg acaacgaagc ggtgcgcgga ctgatccgcc 1740

agcaccgtcc gctcgtgctc atggggtcga tcaacgagaa gatttacctt gcggaactga 1800

aagccggtca cggcccggca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

cgggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggcttgc cctcgatgcc ggtgccgaga gggtcgtgag 2160

agagactgtc gaagccctga cgtgggccgg gcttcggcga gaggaaggga gagaaccaat 2220

gagcgatcat gccgtcaaca cgccggtcca tgccgccagg gcccacgggc accgagcacc 2280

tcgtgccgag ttctacgtct acttcgccgt cattctgctg ggcgccttcc cggtggcctt 2340

cgtgagctgg atcgtctcga cgatccgcca ccgcaggctt cccaagcgcg gccccttcgc 2400

gtccgcctgg ttcgatgcca aggcgatcac gccgctgatt ttccgcgcct gaccgcaggt 2460

caggttgcga cacgccattc gtcgtctccc caaggggcgg cggattaatc gggagggcat 2520

ggtgccttac cgtaacccac gccaccagta ggcaggagga acacgtgtcg tccggtggtc 2580

accaccacca ccaccaccac taataggccg gccctccgtc gcgggcggca cccacgcccg 2640

catcgattcc aaggttcagc cattgagacg gctccgcttc gcgcgcaagc gcgggttggg 2700

ccgactgcaa gcggagaggg aagcatggca ctgctcagct tcgagcgaaa atatcgcgtg 2760

ccgggggggca cgctggtcgg cggaaacctg ttcgacttct gggtcggccc tttctatgtc 2820

ggcttcttcg gggttgcgac gtttttcttc gcggccctgg gtatcattct gattgcctgg 2880

agtgccgtac tccagggtac ctggaacccc caactcatct ctgtctaccc gccggccctt 2940

gaatatggcc tgggaggtgc acccctcgca aaaggcgggc tgtggcagat catcacgatc 3000

tgcgccactg gtgccttcgt cagctgggcg ctgcgcgaag tcgaaatctg ccgtaagctg 3060

ggcatcgggt accacatccc gttcgccttc gcgttcgcca tcctggccta cctgacgctg 3120

gtgctgttcc gcccggtgat gatgggcgcc tggggctatg ccttccccta cgggatctgg 3180

acgcacctcg actgggtgtc gaacacgggc tacacctacg gcaacttcca ctacaaccct 3240

gcccacatga tcgccatctc gttcttcttc acgaacgcgc tggctctggc gctgcacggc 3300

gcccttgtgc tctccgcggc caaccccgag aagggcaagg aaatgcggac gccggatcac 3360

gaggatacgt tcttccgcga tctggtcggc tactcgatcg ggacgctcgg catccaccgc 3420

ctcggcctgc tgctctcgct gagcgccgtc ttcttcagcg ccctctgcat gatcattacc 3480

ggcaccatct ggttcgatca gtgggtcgac tggtggcaat ggtgggtgaa gctgccgtgg 3540

tgggcgaaca tcccgggagg catcaatggc tgagtatcag aacatcttct cccaggtcca 3600

ggtccgcgga ccggccgacc tggggatgac cgaagacgtc aacctggcca accgttcggg 3660

cgtcggtccc ttctcgaccc tgctcggctg gttcggcaac gcccagctcg gcccgatcta 3720

tctcggctcg ctcggcgtcc tgtccctctt ctcgggcctg atgtggttct tcaccatcgg 3780

gatctggttc tggtatcagg cgggctggaa cccggccgtc ttcctgcgcg acctgttctt 3840

-continued

cttctcgctc gagccgccgg cacccgaata cggtctgtcc ttcgcggctc cgctgaagga 3900

aggcgggctg tggctgatcg cgtcgttctt catgttcgtc gcggtctggt cctggtgggg 3960

ccgcacctat ctccgcgctc aggcgctggg catgggcaag cacaccgcct gggcgttcct 4020

ctcggccatc tggctgtgga tggtgctggg cttcatccgt ccgatcctca tggggtcctg 4080

gtcggaagcg gttccctacg gcatcttctc gcacctcgac tggacgaaca acttctcgct 4140

cgtccacggc aacctgttct acaacccctt ccacggtctc tcgatcgcct tcctctacgg 4200

gtcggccctg ctcttcgcga tgcacggtgc gaccatcctc gcggtctccc gcttcggcgg 4260

cgagcgcgag ctggagcaga tcgccgaccg cgggacggca gcggagcggg ccgccctctt 4320

ctggcgctgg accatgggtt tcaacgccac gatggaaggc atccaccgct gggccatctg 4380

gatggcggtc ctcgtgaccc tcaccggcgg catcgggatc ctgctctcgg gcacggtcgt 4440

ggacaactgg tacgtctggg gccagaacca cggcatggcg ccgctgaact gaggagcgat 4500

cacaatggct gacaagacca tcttcaacga tcacctcaac accaatccga agaccaacct 4560

tcgcctctgg gtcgctttcc agatgatgaa gggtgcgggc tgggctggcg gcgtgttctt 4620

cgggacgctc cttctcatcg ggttcttccg ggtggtcggg cggatgcttc cgatccagga 4680

gaaccaggct ccggcgccga acatcaccgg cgctctggag accgggatcg agctgatcaa 4740

gcatctcgtc tgagacaagt ctcggggcag ggcggcgcga ggccgcccgc tcctccaagt 4800

ccgggccata tcgccggccc gggtccgggg cgacaccaca gcccggttcc cttcctgttg 4860

gcgacaggga cctggtgccg tgtggaagac cgcacggcac ccttttgaca ttcacgggag 4920

gctctgatga ccaatcccac cccgcgaccc gaaaccccgc ttttggatcg cgtctgctgc 4980

ccggccgaca tgaaggcgct gagtgacgcc gaactggagc ggctggccga cgaagtgcgt 5040

tccgaggtca gtgatagggg tagtttctta ttttaggcag tttatatgaa attaagacat 5100

gcagatgtca cagtggatat tgaactggtc tcgaaagctc aatatccccc aaagcacaag 5160

cacaaacttc gacatcatgc agaagcgttt cccgaaccgc gtcttcgacg tgggcatcgc 5220

cgagcagcat gccgtgacct tcgcggccgg cctcgccggg gccgggatga agcccttctg 5280

cgcgatctat tcctcgttcc tgcaacgggg ttacgaccag atcgcccatg acgtggcgct 5340

gcagaacctt cccgtccgct tcgtgatcga ccgggcgggg ctcgtggggg ccgatggcgc 5400

gacccatgcg ggggccttcg acgttggctt catcacttcg ctgcccaaca tgaccgtgat 5460

ggccgcggcc gacgaggccg agctcatcca catgatcgcc accgccgtgg ccttcgacga 5520

gggccccatc gccttccgct tcccgcgggg cgaggggtg ggcgtcgaga tgcccgagcg 5580

cgggacggtg ctggagcccg gccggggccg cgtggtgcgc gaagggacgg atgtcgcgat 5640

cctctccttc ggcgcgcatc tgcacgaggc cttgcaggcg gcgaaacttc tcgaggccga 5700

gggggtgagc gtgaccgtgg ccgacgcccg cttctcgcgc ccgctcgaca cggggctcat 5760

cgaccagctc gtgcgccatc acgcggcgct ggtaacggtg gagcaggggg ccatgggcgg 5820

cttcggcgcc catgtcatgc actatctcgc caattccggc ggcttcgacg ggggcctcgc 5880

gctccgggtc atgacgctgc ccgaccgctt catcgagcag gcgagccccg aggacatgta 5940

tgccgatgcg gggctgcggg ccgaggatat caagcttggc gtaatcatgg tcatagctgt 6000

ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa 6060

agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac 6120

-continued

tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg 6180

cggggagagg cggtttgcgt attgggcgct cggtcttgcc ttgctcgtcg gtgatgtact 6240

tcaccagctc cgcgaagtcg ctcttcttga tggagcgcat ggggacgtgc ttggcaatca 6300

cgcgcacccc ccggccgttt tagcggctaa aaaagtcatg gctctgccct cgggcggacc 6360

acgcccatca tgaccttgcc aagctcgtcc tgcttctctt cgatcttcgc cagcagggcg 6420

aggatcgtgg catcaccgaa ccgcgccgtg cgcgggtcgt cggtgagcca gagtttcagc 6480

aggccgccca ggcggcccag gtcgccattg atgcgggcca gctcgcggac gtgctcatag 6540

tccacgacgc ccgtgatttt gtagccctgg ccgacggcca gcaggtaggc cgacaggctc 6600

atgccggccg ccgccgcctt ttcctcaatc gctcttcgtt cgtctggaag gcagtacacc 6660

ttgataggtg ggctgccctt cctggttggc ttggtttcat cagccatccg cttgccctca 6720

tctgttacgc cggcggtagc cggccagcct cgcagagcag gattcccgtt gagcaccgcc 6780

aggtgcgaat aagggacagt gaagaaggaa cacccgctcg cgggtgggcc tacttcacct 6840

atcctgcccg gctgacgccg ttggatacac caaggaaagt ctacacgaac cctttggcaa 6900

aatcctgtat atcgtgcgaa aaaggatgga tataccgaaa aaatcgctat aatgaccccg 6960

aagcagggtt atgcagcgga aaagcgccac gcttcccgaa gggagaaagg cggacaggta 7020

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 7080

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 7140

atgctcgtca gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 7200

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 7260

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 7320

gcgcagcgag tcagtgagcg aggaagcgga agagcgccag aaggccgcca gagaggccga 7380

gcgcggccgt gaggcttgga cgctagggca gggcatgaaa aagcccgtag cgggctgcta 7440

cgggcgtctg acgcggtgga aaggggggagg ggatgttgtc tacatggctc tgctgtagtg 7500

agtgggttgc gctccggcag cggtcctgat caatcgtcac cctttctcgg tccttcaacg 7560

ttcctgacaa cgagcctcct tttcgccaat ccatcgacaa tcaccgcgag tccctgctcg 7620

aacgctgcgt ccggaccggc ttcgtcgaag gcgtctatcg cggcccgcaa cagcggcgag 7680

agcggagcct gttcaacggt gccgccgcgc tcgccggcat cgctgtcgcc ggcctgctcc 7740

tcaagcacgg ccccaacagt gaagtagctg attgtcatca gcgcattgac ggcgtccccg 7800

gccgaaaaac ccgcctcgca gaggaagcga agctgcgcgt cggccgtttc catctgcggt 7860

gcgcccggtc gcgtgccggc atggatgcgc gcgccatcgc ggtaggcgag cagcgcctgc 7920

ctgaagctgc gggcattccc gatcagaaat gagcgccagt cgtcgtcggc tctcggcacc 7980

gaatgcgtat gattctccgc cagcatggct tcggccagtg cgtcgagcag cgcccgcttg 8040

ttcctgaagt gccagtaaag cgccggctgc tgaaccccca accgttccgc cagtttgcgt 8100

gtcgtcagac cgtctacgcc gacctcgttc aacaggtcca gggcggcacg gatcactgta 8160

ttcggctgca actttgtcat gcttgacact ttatcactga taaacataat atgtccacca 8220

acttatcagt gataaagaat ccgcgcgttc aatcggacca gcggaggctg gtccggaggc 8280

cagacatgaa acccaacata cccctgatcg taattctgag cactgtcgcg ctcgacgctg 8340

tcggcatcgg cctgattatg ccggtgctgc cgggcctcct gcgcgatctg gttcactcga 8400

-continued

acgacgtcac cgcccactat ggcattctgc tggcgctgta tgcgttggtg caatttgcct 8460

gcgcacctgt gctgggcgcg ctgtcggatc gtttcgggcg gcggccaatc ttgctcgtct 8520

cgctggccgg cgccactgtc gactacgcca tcatggcgac agcgcctttc ctttgggttc 8580

tctatatcgg gcggatcgtg gccggcatca ccggggcgac tggggcggta gccggcgctt 8640

atattgccga tatcactgat ggcgatgagc gcgcgcggca cttcggcttc atgagcgcct 8700

gtttcgggtt cgggatggtc gcgggacctg tgctcggtgg gctgatgggc ggtttctccc 8760

cccacgctcc gttcttcgcc gcggcagcct tgaacggcct caatttcctg acgggctgtt 8820

tccttttgcc ggagtcgcac aaaggcgaac gccggccgtt acgccgggag gctctcaacc 8880

cgctcgcttc gttccggtgg gcccggggca tgaccgtcgt cgccgccctg atggcggtct 8940

tcttcatcat gcaacttgtc ggacaggtgc cggccgcgct ttgggtcatt ttcggcgagg 9000

atcgctttca ctgggacgcg accacgatcg gcatttcgct tgccgcattt ggcattctgc 9060

attcactcgc ccaggcaatg atcaccggcc ctgtagccgc ccggctcggc gaaaggcggg 9120

cactcatgct cggaatgatt gccgacggca caggctacat cctgcttgcc ttcgcgacac 9180

ggggatggat ggcgttcccg atcatggtcc tgcttgcttc gggtggcatc ggaatgccgg 9240

cgctgcaagc aatgttgtcc aggcaggtgg atgaggaacg tcaggggcag ctgcaaggct 9300

cactggcggc gctcaccagc ctgacctcga tcgtcggacc cctcctcttc acggcgatct 9360

atgcggcttc tataacaacg tggaacgggt gggcatggat tgcaggcgct gccctctact 9420

tgctctgcct gccggcgctg cgtcgcgggc tttggagcgg cgcagggcaa cgagccgatc 9480

gctgatcgtg gaaacgatag gcctatgcca tgcgggtcaa ggcgacttcc ggcaagctat 9540

acgcgcccta ggagtgcggt tggaacgttg gcccagccag atactcccga tcacgagcag 9600

gacgccgatg atttgaagcg cactcagcgt ctgatccaag aacaaccatc ctagcaacac 9660

ggcggtcccc gggctgagaa agcccagtaa ggaaacaact gtaggttcga gtcgcgagat 9720

cccccggaac caaaggaagt aggttaaacc cgctccgatc aggccgagcc acgccaggcc 9780

gagaacattg gttcctgtag gcatcgggat tggcggatca aacactaaag ctactggaac 9840

gagcagaagt cctccggccg ccagttgcca ggcggtaaag gtgagcagag gcacgggagg 9900

ttgccacttg cgggtcagca cggttccgaa cgccatggaa accgcccccg ccaggcccgc 9960

tgcgacgccg acaggatcta gcgctgcgtt tggtgtcaac accaacagcg ccacgcccgc 10020

agttccgcaa atagccccca ggaccgccat caatcgtatc gggctaccta gcagagcggc 10080

agagatgaac acgaccatca gcggctgcac agcgcctacc gtcgccgcga ccccgcccgg 10140

caggcggtag accgaaataa acaacaagct ccagaatagc gaaatattaa gtgcgccgag 10200

gatgaagatg cgcatccacc agattcccgt tggaatctgt cggacgatca tcacgagcaa 10260

taaacccgcc ggcaacgccc gcagcagcat accggcgacc cctcggcctc gctgttcggg 10320

ctccacgaaa acgccggaca gatgcgcctt gtgagcgtcc ttggggccgt cctcctgttt 10380

gaagaccgac agcccaatga tctcgccgtc gatgtaggcg ccgaatgcca 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

gcagtgcccg cttgttcctg aaatgccagt aaagcgctgg ctgctgaacc cccagccgga 10740

actgacccca caaggcccta gcgtttgcaa tgcaccaggt catcattgac ccaggcgtgt 10800

tccaccaggc cgctgcctcg caactcttcg caggcttcgc cgacctgctc gcgccacttc 10860

ttcacgcggg tggaatccga tccgcacatg aggcggaagg tttccagctt gagcgggtac 10920

ggctcccggt gcgagctgaa atagtcgaac atccgtcggg ccgtcggcga cagcttgcgg 10980

tacttctccc atatgaattt cgtgtagtgg tcgccagcaa acagcacgac gatttcctcg 11040

tcgatcagga cctggcaacg ggacgttttc ttgccacggt ccaggacgcg gaagcggtgc 11100

agcagcgaca ccgattccag gtgcccaacg cggtcggacg tgaagcccat cgccgtcgcc 11160

tgtaggcgcg acaggcattc ctcggccttc gtgtaatacc ggccattgat cgaccagccc 11220

aggtcctggc aaagctcgta gaacgtgaag gtgatcggct cgccgatagg ggtgcgcttc 11280

gcgtactcca acacctgctg ccacaccagt tcgtcatcgt cggcccgcag ctcgacgccg 11340

gtgtaggtga tcttcacgtc cttgttgacg tggaaaatga ccttgttttg cagcgcctcg 11400

cgcgggattt tcttgttgcg cgtggtgaac agggcagagc gggccgtgtc gtttggcatc 11460

gctcgcatcg tgtccggcca cggcgcaata tcgaacaagg aaagctgcat ttccttgatc 11520

tgctgcttcg tgtgtttcag caacgcggcc tgcttggcct cgctgacctg ttttgccagg 11580

tcctcgccgg cggtttttcg cttcttggtc gtcatagttc ctcgcgtgtc gatggtcatc 11640

gacttcgcca aacctgccgc ctcctgttcg agacgacgcg aacgctccac ggcggccgat 11700

ggcgcgggca gggcaggggg agccagttgc acgctgtcgc gctcgatctt ggccgtagct 11760

tgctggacca tcgagccgac ggactggaag gtttcgcggg gcgcacgcat gacggtgcgg 11820

cttgcgatgg tttcggcatc ctcggcggaa aaccccgcgt cgatcagttc ttgcctgtat 11880

gccttccggt caaacgtccg attcattcac cctccttgcg ggattgcccc gactcacgcc 11940

ggggcaatgt gcccttattc ctgatttgac ccgcctggtg ccttggtgtc cagataatcc 12000

accttatcgg caatgaagtc ggtcccgtag accgtctggc cgtccttctc gtacttggta 12060

ttccgaatct tgccctgcac gaataccagc gacccccttgc ccaaatactt gccgtgggcc 12120

tcggcctgag agccaaaaca cttgatgcgg aagaagtcgg tgcgctcctg cttgtcgccg 12180

gtcgtggccg cgccaacctt tgcgatccgc aagcgcgcgg tcgccatctt cacgctggaa 12240

cagtacgtcg aggcgggcat catgacccgc gagcaatacg aggtcattaa aagcgccgtg 12300

attgatgata tagcggcccg gctgctcctg gttctcgcgc accgaaatgg gtgacttcac 12360

cccgcgctct ttgatcgtgg caccgatttc cgcgatgctc tccggggaaa agccggggtt 12420

gtcggccgtc cgcggctgat gcggatcttc gtcgatcagg tccaggtcca gctcgatagg 12480

gccggaaccg ccctgagacg ccgcaggagc gtccaggagg ctcgacaggt cgccgatgct 12540

atccaacccc aggccggacg gctgcgccgc gcctgcggct tcctgagcgg ccgcagcggt 12600

gtttttcttg gtggtcttgg cttgagccgc agtcattggg aaatctccat cttcgtgaac 12660

acgtaatcag ccagggcgcg aacctctttc gatgccttgc gcgcggccgt tttcttgatc 12720

ttccagaccg gcacaccgga tgcgagggca tcggcgatgc tgctgcgcag gccaacggtg 12780

gccggaatca tcatcttggg gtacgcggcc agcagctcgg cttggtggcg cgcgtggcgc 12840

ggattccgcg catcgacctt gctgggcacc atgccaagga attgcagctt ggcgttcttc 12900

tggcgcacgt tcgcaatggt cgtgaccatc ttcttgatgc cctggatgct gtacgcctca 12960

-continued

agctcgatgg gggacagcac atagtcggcc gcgaagaggg cggccgccag gccgacgcca 13020

agggtcgggg ccgtgtcgat caggcacacg tcgaagcctt ggttcgccag ggccttgatg 13080

ttcgccccga acagctcgcg ggcgtcgtcc agcgacagcc gttcggcgtt cgccagtacc 13140

gggttggact cgatgagggc gaggcgcgcg gcctggccgt cgccggctgc gggtgcggtt 13200

tcggtccagc cgccggcagg gacagcgccg aacagcttgc ttgcatgcag gccggtagca 13260

aagtccttga gcgtgtagga cgcattgccc tgggggtcca ggtcgatcac ggcaacccgc 13320

aagccgcgct cgaaaaagtc gaaggcaaga tgcacaaggg tcgaagtctt gccgacgccg 13380

cctttctggt tggccgtgac caaagttttc atcgtttggt ttcctgtttt ttcttggcgt 13440

ccgcttccca cttccggacg atgtacgcct gatgttccgg cagaaccgcc gttacccgcg 13500

cgtacccctc gggcaagttc ttgtcctcga acgcggccca cacgcgatgc accgcttgcg 13560

acactgcgcc cctggtcagt cccagcgacg ttgcgaacgt cgcctgtggc ttcccatcga 13620

ctaagacgcc ccgcgctatc tcgatggtct gctgccccac ttccagcccc tggatcgcct 13680

cctggaactg gctttcggta agccgtttct tcatggataa cacccataat ttgctccgcg 13740

ccttggttga acatagcggt gacagccgcc agcacatgag agaagtttag ctaaacattt 13800

ctcgcacgtc aacaccttta gccgctaaaa ctcgtccttg gcgtaacaaa acaaaagccc 13860

ggaaaccggg ctttcgtctc ttgccgctta tggctctgca cccggctcca tcaccaacag 13920

gtcgcgcacg cgcttcactc ggttgcggat cgacactgcc agcccaacaa agccggttgc 13980

cgccgccgcc aggatcgcgc cgatgatgcc ggccacaccg gccatcgccc accaggtcgc 14040

cgccttccgg ttccattcct gctggtactg cttcgcaatg ctggacctcg gctcaccata 14100

ggctgaccgc tcgatggcgt atgccgcttc tccccttggc gtaaaaccca gcgccgcagg 14160

cggcattgcc atgctgcccg ccgctttccc gaccacgacg cgcgcaccag gcttgcggtc 14220

cagaccttcg gccacggcga gctgcgcaag gacataatca gccgccgact tggctccacg 14280

cgcctcgatc agctcttgca ctcgcgcgaa atccttggcc tccacggccg ccatgaatcg 14340

cgcacgcggc gaaggctccg cagggccg 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

-continued

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggcgactag ttccatatga tagatctcac 2580

caccaccacc accaccacca ccaccactaa taggccggcc ctccgtcgcg ggcggcaccc 2640

acgcccgcat cgattccaag gttcagccat tgagacggct ccgcttcgcg cgcaagcgcg 2700

ggttgggccg actgcaagcg gagagggaag catggcactg ctcagcttcg agcgaaaata 2760

tcgcgtgccg gggggcacgc tggtcggcgg aaacctgttc gacttctggg tcggcccttt 2820

-continued

ctatgtcggc ttcttcgggg ttgcgacgtt tttcttcgcg gccctgggta tcattctgat 2880

tgcctggagt gccgtactcc agggtacctg gaacccccaa ctcatctctg tctacccgcc 2940

ggcccttgaa tatggcctgg gaggtgcacc cctcgcaaaa ggcgggctgt ggcagatcat 3000

cacgatctgc gccactggtg ccttcgtcag ctgggcgctg cgcgaagtcg aaatctgccg 3060

taagctgggc atcgggtacc acatcccgtt cgccttcgcg ttcgccatcc tggcctacct 3120

gacgctggtg ctgttccgcc cggtgatgat gggcgcctgg ggctatgcct tcccctacgg 3180

gatctggacg cacctcgact gggtgtcgaa cacgggctac acctacggca acttccacta 3240

caaccctgcc cacatgatcg ccatctcgtt cttcttcacg aacgcgctgg ctctggcgct 3300

gcacggcgcc cttgtgctct ccgcggccaa ccccgagaag ggcaaggaaa tgcggacgcc 3360

ggatcacgag gatacgttct tccgcgatct ggtcggctac tcgatcggga cgctcggcat 3420

ccaccgcctc ggcctgctgc tctcgctgag cgccgtcttc ttcagcgccc tctgcatgat 3480

cattaccggc accatctggt tcgatcagtg ggtcgactgg tggcaatggt gggtgaagct 3540

gccgtggtgg gcgaacatcc cgggaggcat caatggctga gtatcagaac atcttctccc 3600

aggtccaggt ccgcggaccg gccgacctgg ggatgaccga agacgtcaac ctggccaacc 3660

gttcgggcgt cggtcccttc tcgaccctgc tcggctggtt cggcaacgcc cagctcggcc 3720

cgatctatct cggctcgctc ggcgtcctgt ccctcttctc gggcctgatg tggttcttca 3780

ccatcgggat ctggttctgg tatcaggcgg gctggaaccc ggccgtcttc ctgcgcgacc 3840

tgttcttctt ctcgctcgag ccgccggcac ccgaatacgg tctgtccttc gcggctccgc 3900

tgaaggaagg cgggctgtgg ctgatcgcgt cgttcttcat gttcgtcgcg gtctggtcct 3960

ggtggggccg cacctatctc cgcgctcagg cgctgggcat gggcaagcac accgcctggg 4020

cgttcctctc ggccatctgg ctgtggatgg tgctgggctt catccgtccg atcctcatgg 4080

ggtcctggtc ggaagcggtt ccctacggca tcttctcgca cctcgactgg acgaacaact 4140

tctcgctcgt ccacggcaac ctgttctaca acccttcca cggtctctcg atcgccttcc 4200

tctacgggtc ggccctgctc ttcgcgatgc acggtgcgac catcctcgcg gtctcccgct 4260

tcggcggcga gcgcgagctg gagcagatcg ccgaccgcgg gacggcagcg gagcgggccg 4320

ccctcttctg gcgctggacc atgggtttca acgccacgat ggaaggcatc caccgctggg 4380

ccatctggat ggcggtcctc gtgaccctca ccggcggcat cgggatcctg ctctcgggca 4440

cggtcgtgga caactggtac gtctggggcc agaaccacgg catggcgccg ctgaactgag 4500

gagcgatcac aatggctgac aagaccatct tcaacgatca cctcaacacc aatccgaaga 4560

ccaaccttcg cctctgggtc gctttccaga tgatgaaggg tgcgggctgg gctggcggcg 4620

tgttcttcgg gacgctcctt ctcatcgggt tcttccgggt ggtcgggcgg atgcttccga 4680

tccaggagaa ccaggctccg gcgccgaaca tcaccggcgc tctggagacc gggatcgagc 4740

tgatcaagca tctcgtctga gacaagtctc ggggcagggc ggcgcgaggc cgcccgctcc 4800

tccaagtccg ggccatatcg ccggcccggg tccggggcga caccacagcc cggttccctt 4860

cctgttggcg acagggacct ggtgccgtgt ggaagaccgc acggcaccct tttgacattc 4920

acgggaggct ctgatgacca atcccacccc gcgacccgaa accccgcttt tggatcgcgt 4980

ctgctgcccg gccgacatga aggcgctgag tgacgccgaa ctggagcggc tggccgacga 5040

agtgcgttcc gaggtcagtg ataggggtag tttcttattt taggcagttt atatgaaatt 5100

-continued

aagacatgca gatgtcacag tggatattga actggtctcg aaagctcaat atcccccaaa 5160

gcacaagcac aaacttcgac atcatgcaga agcgtttccc gaaccgcgtc ttcgacgtgg 5220

gcatcgccga gcagcatgcc gtgaccttcg cggccggcct cgccggggcc gggatgaagc 5280

ccttctgcgc gatctattcc tcgttcctgc aacggggtta cgaccagatc gcccatgacg 5340

tggcgctgca gaaccttccc gtccgcttcg tgatcgaccg ggcggggctc gtgggggccg 5400

atggcgcgac ccatgcgggg gccttcgacg ttggcttcat cacttcgctg cccaacatga 5460

ccgtgatggc cgcggccgac gaggccgagc tcatccacat gatcgccacc gccgtggcct 5520

tcgacgaggg ccccatcgcc ttccgcttcc cgcggggcga gggggtgggc gtcgagatgc 5580

ccgagcgcgg gacggtgctg gagcccggcc ggggccgcgt ggtgcgcgaa gggacggatg 5640

tcgcgatcct ctccttcggc gcgcatctgc acgaggcctt gcaggcggcg aaacttctcg 5700

aggccgaggg ggtgagcgtg accgtggccg acgcccgctt ctcgcgcccg ctcgacacgg 5760

ggctcatcga ccagctcgtg cgccatcacg cggcgctggt aacggtggag caggggggcca 5820

tgggcggctt cggcgcccat gtcatgcact atctcgccaa ttccggcggc ttcgacgggg 5880

gcctcgcgct ccgggtcatg acgctgcccg accgcttcat cgagcaggcg agccccgagg 5940

acatgtatgc cgatgcgggg ctgcgggccg aggatatcaa gcttggcgta atcatggtca 6000

tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat acgagccgga 6060

agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt aattgcgttg 6120

cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc 6180

caacgcgcgg ggagaggcgg tttgcgtatt gggcgctcgg tcttgccttg ctcgtcggtg 6240

atgtacttca ccagctccgc gaagtcgctc ttcttgatgg agcgcatggg gacgtgcttg 6300

gcaatcacgc gcacccccg gccgttttag cggctaaaaa agtcatggct ctgccctcgg 6360

gcggaccacg cccatcatga ccttgccaag ctcgtcctgc ttctcttcga tcttcgccag 6420

cagggcgagg atcgtggcat caccgaaccg cgccgtgcgc gggtcgtcgg tgagccagag 6480

tttcagcagg ccgcccaggc ggcccaggtc gccattgatg cgggccagct cgcggacgtg 6540

ctcatagtcc acgacgcccg tgattttgta gccctggccg acggccagca ggtaggccga 6600

caggctcatg ccggccgccg ccgccttttc ctcaatcgct cttcgttcgt ctggaaggca 6660

gtacaccttg ataggtgggc tgcccttcct ggttggcttg gtttcatcag ccatccgctt 6720

gccctcatct gttacgccgg cggtagccgg ccagcctcgc agagcaggat tcccgttgag 6780

caccgccagg tgcgaataag ggacagtgaa gaaggaacac ccgctcgcgg gtgggcctac 6840

ttcacctatc ctgcccggct gacgccgttg gatacaccaa ggaaagtcta cacgaaccct 6900

ttggcaaaat cctgtatatc gtgcgaaaaa ggatggatat accgaaaaaa tcgctataat 6960

gaccccgaag cagggttatg cagcggaaaa gcgccacgct tcccgaaggg agaaaggcgg 7020

acaggtatcc ggtaagcggc agggtcggaa caggagagcg cacgagggag cttccagggg 7080

gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca cctctgactt gagcgtcgat 7140

ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa cgccagcaac gcggcctttt 7200

tacggttcct ggccttttgc tggccttttg ctcacatgtt ctttcctgcg ttatcccctg 7260

attctgtgga taaccgtatt accgcctttg agtgagctga taccgctcgc cgcagccgaa 7320

cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga gcgccagaag gccgccagag 7380

-continued

aggccgagcg cggccgtgag gcttggacgc tagggcaggg catgaaaaag cccgtagcgg 7440

gctgctacgg gcgtctgacg cggtggaaag ggggagggga tgttgtctac atggctctgc 7500

tgtagtgagt gggttgcgct ccggcagcgg tcctgatcaa tcgtcaccct ttctcggtcc 7560

ttcaacgttc ctgacaacga gcctcctttt cgccaatcca tcgacaatca ccgcgagtcc 7620

ctgctcgaac gctgcgtccg gaccggcttc gtcgaaggcg tctatcgcgg cccgcaacag 7680

cggcgagagc ggagcctgtt caacggtgcc gccgcgctcg ccggcatcgc tgtcgccggc 7740

ctgctcctca agcacggccc caacagtgaa gtagctgatt gtcatcagcg cattgacggc 7800

gtccccggcc gaaaaacccg cctcgcagag gaagcgaagc tgcgcgtcgg ccgtttccat 7860

ctgcggtgcg cccggtcgcg tgccggcatg gatgcgcgcg ccatcgcggt aggcgagcag 7920

cgcctgcctg aagctgcggg cattcccgat cagaaatgag cgccagtcgt cgtcggctct 7980

cggcaccgaa tgcgtatgat tctccgccag catggcttcg gccagtgcgt cgagcagcgc 8040

ccgcttgttc ctgaagtgcc agtaaagcgc cggctgctga acccccaacc gttccgccag 8100

tttgcgtgtc gtcagaccgt ctacgccgac ctcgttcaac aggtccaggg cggcacggat 8160

cactgtattc ggctgcaact ttgtcatgct tgacacttta tcactgataa acataatatg 8220

tccaccaact tatcagtgat aaagaatccg cgcgttcaat cggaccagcg gaggctggtc 8280

cggaggccag acatgaaacc caacataccc ctgatcgtaa ttctgagcac tgtcgcgctc 8340

gacgctgtcg gcatcggcct gattatgccg gtgctgccgg gcctcctgcg cgatctggtt 8400

cactcgaacg acgtcaccgc ccactatggc attctgctgg cgctgtatgc gttggtgcaa 8460

tttgcctgcg cacctgtgct gggcgcgctg tcggatcgtt tcgggcggcg gccaatcttg 8520

ctcgtctcgc tggccggcgc cactgtcgac tacgccatca tggcgacagc gcctttcctt 8580

tgggttctct atatcgggcg gatcgtggcc ggcatcaccg gggcgactgg ggcggtagcc 8640

ggcgcttata ttgccgatat cactgatggc gatgagcgcg cgcggcactt cggcttcatg 8700

agcgcctgtt tcgggttcgg gatggtcgcg ggacctgtgc tcggtgggct gatgggcggt 8760

ttctcccccc acgctccgtt cttcgccgcg gcagccttga acggcctcaa tttcctgacg 8820

ggctgtttcc ttttgccgga gtcgcacaaa ggcgaacgcc ggccgttacg ccgggaggct 8880

ctcaacccgc tcgcttcgtt ccggtgggcc cggggcatga ccgtcgtcgc cgccctgatg 8940

gcggtcttct tcatcatgca acttgtcgga caggtgccgg ccgcgctttg ggtcattttc 9000

ggcgaggatc gctttcactg ggacgcgacc acgatcggca tttcgcttgc cgcatttggc 9060

attctgcatt cactcgccca ggcaatgatc accggccctg tagccgcccg 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 tcggacccct cctcttcacg 9360

gcgatctatg cggcttctat aacaacgtgg aacgggtggg catggattgc aggcgctgcc 9420

ctctacttgc tctgcctgcc ggcgctgcgt cgcgggcttt ggagcggcgc agggcaacga 9480

gccgatcgct gatcgtggaa acgataggcc tatgccatgc gggtcaaggc gacttccggc 9540

aagctatacg cgccctagga gtgcggttgg aacgttggcc cagccagata ctcccgatca 9600

cgagcaggac gccgatgatt tgaagcgcac tcagcgtctg atccaagaac aaccatccta 9660

-continued

gcaacacggc ggtccccggg ctgagaaagc ccagtaagga aacaactgta ggttcgagtc 9720

gcgagatccc ccggaaccaa aggaagtagg ttaaacccgc tccgatcagg ccgagccacg 9780

ccaggccgag aacattggtt cctgtaggca tcgggattgg cggatcaaac actaaagcta 9840

ctggaacgag cagaagtcct ccggccgcca gttgccaggc ggtaaaggtg agcagaggca 9900

cgggaggttg ccacttgcgg gtcagcacgg ttccgaacgc catggaaacc gcccccgcca 9960

ggcccgctgc gacgccgaca ggatctagcg ctgcgtttgg tgtcaacacc aacagcgcca 10020

cgcccgcagt tccgcaaata gcccccagga ccgccatcaa tcgtatcggg ctacctagca 10080

gagcggcaga gatgaacacg accatcagcg gctgcacagc gcctaccgtc gccgcgaccc 10140

cgcccggcag gcggtagacc gaaataaaca acaagctcca gaatagcgaa atattaagtg 10200

cgccgaggat gaagatgcgc atccaccaga ttcccgttgg aatctgtcgg acgatcatca 10260

cgagcaataa acccgccggc aacgcccgca gcagcatacc ggcgacccct cggcctcgct 10320

gttcgggctc cacgaaaacg ccggacagat gcgccttgtg agcgtccttg gggccgtcct 10380

cctgtttgaa gaccgacagc ccaatgatct cgccgtcgat gtaggcgccg aatgccacgg 10440

catctcgcaa ccgttcagcg aacgcctcca tgggcttttt ctcctcgtgc tcgtaaacgg 10500

acccgaacat ctctggagct ttcttcaggg ccgacaatcg gatctcgcgg aaatcctgca 10560

cgtcggccgc tccaagccgt cgaatctgag ccttaatcac aattgtcaat tttaatcctc 10620

tgtttatcgg cagttcgtag agcgcgccgt gcgtcccgag cgatactgag cgaagcaagt 10680

gcgtcgagca gtgcccgctt gttcctgaaa tgccagtaaa gcgctggctg ctgaaccccc 10740

agccggaact gaccccacaa ggccctagcg tttgcaatgc accaggtcat cattgaccca 10800

ggcgtgttcc accaggccgc tgcctcgcaa ctcttcgcag gcttcgccga cctgctcgcg 10860

ccacttcttc acgcgggtgg aatccgatcc gcacatgagg cggaaggttt ccagcttgag 10920

cgggtacggc tcccggtgcg agctgaaata gtcgaacatc cgtcgggccg tcggcgacag 10980

cttgcggtac ttctcccata tgaatttcgt gtagtggtcg ccagcaaaca gcacgacgat 11040

ttcctcgtcg atcaggacct ggcaacggga cgttttcttg ccacggtcca ggacgcggaa 11100

gcggtgcagc agcgacaccg attccaggtg cccaacgcgg tcggacgtga agcccatcgc 11160

cgtcgcctgt aggcgcgaca ggcattcctc ggccttcgtg taataccggc cattgatcga 11220

ccagcccagg tcctggcaaa gctcgtagaa cgtgaaggtg atcggctcgc cgataggggt 11280

gcgcttcgcg tactccaaca cctgctgcca caccagttcg tcatcgtcgg cccgcagctc 11340

gacgccggtg taggtgatct tcacgtcctt gttgacgtgg aaaatgacct tgttttgcag 11400

cgcctcgcgc gggattttct tgttgcgcgt ggtgaacagg gcagagcggg ccgtgtcgtt 11460

tggcatcgct cgcatcgtgt ccggccacgg cgcaatatcg aacaaggaaa gctgcatttc 11520

cttgatctgc tgcttcgtgt gtttcagcaa cgcggcctgc ttggcctcgc tgacctgttt 11580

tgccaggtcc tcgccggcgg tttttcgctt cttggtcgtc atagttcctc gcgtgtcgat 11640

ggtcatcgac ttcgccaaac ctgccgcctc ctgttcgaga cgacgcgaac gctccacggc 11700

ggccgatggc gcgggcaggg caggggggagc cagttgcacg ctgtcgcgct cgatcttggc 11760

cgtagcttgc tggaccatcg agccgacgga ctggaaggtt tcgcggggcg cacgcatgac 11820

ggtgcggctt gcgatggttt cggcatcctc ggcggaaaac cccgcgtcga tcagttcttg 11880

cctgtatgcc ttccggtcaa acgtccgatt cattcaccct ccttgcggga ttgccccgac 11940

-continued

tcacgccggg gcaatgtgcc cttattcctg atttgacccg cctggtgcct tggtgtccag 12000

ataatccacc ttatcggcaa tgaagtcggt cccgtagacc gtctggccgt ccttctcgta 12060

cttggtattc cgaatcttgc cctgcacgaa taccagcgac cccttgccca aatacttgcc 12120

gtgggcctcg gcctgagagc caaaacactt gatgcggaag aagtcggtgc gctcctgctt 12180

gtcgccggtc gtggccgcgc caacctttgc gatccgcaag cgcgcggtcg ccatcttcac 12240

gctggaacag tacgtcgagg cgggcatcat gacccgcgag caatacgagg tcattaaaag 12300

cgccgtgatt gatgatatag cggcccggct gctcctggtt ctcgcgcacc gaaatgggtg 12360

acttcacccc gcgctctttg atcgtggcac cgatttccgc gatgctctcc ggggaaaagc 12420

cggggttgtc ggccgtccgc ggctgatgcg gatcttcgtc gatcaggtcc aggtccagct 12480

cgatagggcc ggaaccgccc tgagacgccg caggagcgtc caggaggctc gacaggtcgc 12540

cgatgctatc caaccccagg ccggacggct gcgccgcgcc tgcggcttcc tgagcggccg 12600

cagcggtgtt tttcttggtg gtcttggctt gagccgcagt cattgggaaa tctccatctt 12660

cgtgaacacg taatcagcca gggcgcgaac ctctttcgat gccttgcgcg cggccgtttt 12720

cttgatcttc cagaccggca caccggatgc gagggcatcg gcgatgctgc tgcgcaggcc 12780

aacggtggcc ggaatcatca tcttggggta cgcggccagc agctcggctt ggtggcgcgc 12840

gtggcgcgga ttccgcgcat cgaccttgct gggcaccatg ccaaggaatt gcagcttggc 12900

gttcttctgg cgcacgttcg caatggtcgt gaccatcttc ttgatgccct ggatgctgta 12960

cgcctcaagc tcgatggggg acagcacata gtcggccgcg aagagggcgg ccgccaggcc 13020

gacgccaagg gtcggggccg tgtcgatcag gcacacgtcg aagccttggt tcgccagggc 13080

cttgatgttc gccccgaaca gctcgcgggc gtcgtccagc gacagccgtt cggcgttcgc 13140

cagtaccggg ttggactcga tgagggcgag gcgcgcggcc tggccgtcgc cggctgcggg 13200

tgcggtttcg gtccagccgc cggcagggac agcgccgaac agcttgcttg catgcaggcc 13260

ggtagcaaag tccttgagcg tgtaggacgc attgccctgg gggtccaggt cgatcacggc 13320

aacccgcaag ccgcgctcga aaaagtcgaa ggcaagatgc acaagggtcg aagtcttgcc 13380

gacgccgcct ttctggttgg ccgtgaccaa agttttcatc gtttggtttc ctgtttttc 13440

ttggcgtccg cttcccactt ccggacgatg tacgcctgat gttccggcag aaccgccgtt 13500

acccgcgcgt acccctcggg caagttcttg tcctcgaacg cggcccacac gcgatgcacc 13560

gcttgcgaca ctgcgcccct ggtcagtccc agcgacgttg cgaacgtcgc ctgtggcttc 13620

ccatcgacta agacgccccg cgctatctcg atggtctgct gccccacttc cagcccctgg 13680

atcgcctcct ggaactggct ttcggtaagc cgtttcttca tggataacac ccataatttg 13740

ctccgcgcct tggttgaaca tagcggtgac agccgccagc acatgagaga agtttagcta 13800

aacatttctc gcacgtcaac acctttagcc gctaaaactc gtccttggcg taacaaaaca 13860

aaagcccgga aaccgggctt tcgtctcttg ccgcttatgg ctctgcaccc ggctccatca 13920

ccaacaggtc gcgcacgcgc ttcactcggt tgcggatcga cactgccagc ccaacaaagc 13980

cggttgccgc cgccgccagg atcgcgccga tgatgccggc cacaccggcc atcgcccacc 14040

aggtcgccgc cttccggttc cattcctgct ggtactgctt cgcaatgctg gacctcggct 14100

caccataggc tgaccgctcg atggcgtatg ccgcttctcc ccttggcgta aaacccagcg 14160

ccgcaggcgg cattgccatg ctgcccgccg ctttcccgac cacgacgcgc gcaccaggct 14220

-continued

tgcggtccag accttcggcc acggcgagct gcgcaaggac ataatcagcc gccgacttgg 14280

ctccacgcgc ctcgatcagc tcttgcactc gcgcgaaatc cttggcctcc acggccgcca 14340

tgaatcgcgc acgcggcgaa ggctccgcag ggccg 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggag gcaccccccc tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgccccc catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 900

tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 960

aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 1080

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

-continued

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggcgactag ttccatatga tagatctcac 2580

caccaccacc accaccacca ccaccaccac caccactaat aggccggccc tccgtcgcgg 2640

gcggcaccca cgcccgcatc gattccaagg ttcagccatt gagacggctc cgcttcgcgc 2700

gcaagcgcgg gttgggccga ctgcaagcgg agagggaagc atggcactgc tcagcttcga 2760

gcgaaaatat cgcgtgccgg ggggcacgct ggtcggcgga aacctgttcg acttctgggt 2820

cggccctttc tatgtcggct tcttcggggt tgcgacgttt ttcttcgcgg ccctgggtat 2880

cattctgatt gcctggagtg ccgtactcca gggtacctgg aacccccaac tcatctctgt 2940

ctacccgccg gcccttgaat atggcctggg aggtgcaccc ctcgcaaaag gcgggctgtg 3000

gcagatcatc acgatctgcg ccactggtgc cttcgtcagc tgggcgctgc gcgaagtcga 3060

aatctgccgt aagctgggca tcgggtacca catcccgttc gccttcgcgt tcgccatcct 3120

ggcctacctg acgctggtgc tgttccgccc ggtgatgatg ggcgcctggg gctatgcctt 3180

cccctacggg atctggacgc acctcgactg ggtgtcgaac acgggctaca cctacggcaa 3240

cttccactac aaccctgccc acatgatcgc catctcgttc ttcttcacga acgcgctggc 3300

tctggcgctg cacggcgccc ttgtgctctc cgcggccaac cccgagaagg gcaaggaaat 3360

gcggacgccg gatcacgagg atacgttctt ccgcgatctg gtcggctact cgatcgggac 3420

gctcggcatc caccgcctcg gcctgctgct ctcgctgagc gccgtcttct tcagcgccct 3480

ctgcatgatc attaccggca ccatctggtt cgatcagtgg gtcgactggt ggcaatggtg 3540

ggtgaagctg ccgtggtggg cgaacatccc gggaggcatc aatggctgag tatcagaaca 3600

tcttctccca ggtccaggtc cgcggaccgg ccgacctggg gatgaccgaa gacgtcaacc 3660

tggccaaccg ttcgggcgtc ggtcccttct cgaccctgct cggctggttc ggcaacgccc 3720

agctcggccc gatctatctc ggctcgctcg gcgtcctgtc cctcttctcg ggcctgatgt 3780

ggttcttcac catcgggatc tggttctggt atcaggcggg ctggaacccg gccgtcttcc 3840

tgcgcgacct gttcttcttc tcgctcgagc cgccggcacc cgaatacggt ctgtccttcg 3900

cggctccgct gaaggaaggc gggctgtggc tgatcgcgtc gttcttcatg ttcgtcgcgg 3960

tctggtcctg gtggggccgc acctatctcc gcgctcaggc gctgggcatg ggcaagcaca 4020

-continued

ccgcctgggc gttcctctcg gccatctggc tgtggatggt gctgggcttc atccgtccga 4080

tcctcatggg gtcctggtcg gaagcggttc cctacggcat cttctcgcac ctcgactgga 4140

cgaacaactt ctcgctcgtc cacggcaacc tgttctacaa ccccttccac ggtctctcga 4200

tcgccttcct ctacgggtcg gccctgctct tcgcgatgca cggtgcgacc atcctcgcgg 4260

tctcccgctt cggcggcgag cgcgagctgg agcagatcgc cgaccgcggg acggcagcgg 4320

agcgggccgc cctcttctgg cgctggacca tgggtttcaa cgccacgatg gaaggcatcc 4380

accgctgggc catctggatg gcggtcctcg tgaccctcac cggcggcatc gggatcctgc 4440

tctcgggcac ggtcgtggac aactggtacg tctggggcca gaaccacggc atggcgccgc 4500

tgaactgagg agcgatcaca atggctgaca agaccatctt caacgatcac ctcaacacca 4560

atccgaagac caaccttcgc ctctgggtcg ctttccagat gatgaagggt gcgggctggg 4620

ctggcggcgt gttcttcggg acgctccttc tcatcgggtt cttccgggtg gtcgggcgga 4680

tgcttccgat ccaggagaac caggctccgg cgccgaacat caccggcgct ctggagaccg 4740

ggatcgagct gatcaagcat ctcgtctgag acaagtctcg gggcagggcg gcgcgaggcc 4800

gcccgctcct ccaagtccgg gccatatcgc cggcccgggt ccggggcgac accacagccc 4860

ggttcccttc ctgttggcga cagggacctg gtgccgtgtg gaagaccgca cggcaccctt 4920

ttgacattca cgggaggctc tgatgaccaa tcccaccccg cgacccgaaa ccccgctttt 4980

ggatcgcgtc tgctgcccgg ccgacatgaa ggcgctgagt gacgccgaac tggagcggct 5040

ggccgacgaa gtgcgttccg aggtcagtga taggggtagt ttcttatttt aggcagttta 5100

tatgaaatta agacatgcag atgtcacagt ggatattgaa ctggtctcga aagctcaata 5160

tcccccaaag cacaagcaca aacttcgaca tcatgcagaa gcgtttcccg aaccgcgtct 5220

tcgacgtggg catcgccgag cagcatgccg tgaccttcgc ggccggcctc gccggggccg 5280

ggatgaagcc cttctgcgcg atctattcct cgttcctgca acggggttac gaccagatcg 5340

cccatgacgt ggcgctgcag aaccttcccg tccgcttcgt gatcgaccgg gcggggctcg 5400

tgggggccga tggcgcgacc catgcggggg ccttcgacgt tggcttcatc acttcgctgc 5460

ccaacatgac cgtgatggcc gcggccgacg aggccgagct catccacatg atcgccaccg 5520

ccgtggcctt cgacgagggc cccatcgcct tccgcttccc gcggggcgag ggggtgggcg 5580

tcgagatgcc cgagcgcggg acggtgctgg agcccggccg gggccgcgtg gtgcgcgaag 5640

ggacggatgt cgcgatcctc tccttcggcg cgcatctgca cgaggccttg caggcggcga 5700

aacttctcga ggccgagggg gtgagcgtga ccgtggccga cgcccgcttc tcgcgcccgc 5760

tcgacacggg gctcatcgac cagctcgtgc gccatcacgc ggcgctggta acggtggagc 5820

aggggccat gggcggcttc ggcgcccatg tcatgcacta tctcgccaat tccggcggct 5880

tcgacggggg cctcgcgctc cgggtcatga cgctgcccga ccgcttcatc gagcaggcga 5940

gccccgagga catgtatgcc gatgcggggc tgcgggccga ggatatcaag cttggcgtaa 6000

tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata 6060

cgagccggaa gcataaagtg taaagcctgg ggtgcctaat gagtgagcta actcacatta 6120

attgcgttgc gctcactgcc cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa 6180

tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg ggcgctcggt cttgccttgc 6240

tcgtcggtga tgtacttcac cagctccgcg aagtcgctct tcttgatgga gcgcatgggg 6300

-continued

acgtgcttgg caatcacgcg caccccccgg ccgttttagc ggctaaaaaa gtcatggctc 6360

tgccctcggg cggaccacgc ccatcatgac cttgccaagc tcgtcctgct tctcttcgat 6420

cttcgccagc agggcgagga tcgtggcatc accgaaccgc gccgtgcgcg ggtcgtcggt 6480

gagccagagt ttcagcaggc cgcccaggcg gcccaggtcg ccattgatgc gggccagctc 6540

gcggacgtgc tcatagtcca cgacgcccgt gattttgtag ccctggccga cggccagcag 6600

gtaggccgac aggctcatgc cggccgccgc cgcctttccc tcaatcgctc ttcgttcgtc 6660

tggaaggcag tacaccttga taggtgggct gcccttcctg 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 cgccacgctt cccgaaggga 7020

gaaaggcgga caggtatccg gtaagcggca gggtcggaac aggagagcgc acgagggagc 7080

ttccaggggg aaacgcctgg tatctttata gtcctgtcgg gtttcgccac ctctgacttg 7140

agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct atggaaaaac gccagcaacg 7200

cggccttttt acggttcctg gccttttgct ggccttttgc tcacatgttc tttcctgcgt 7260

tatcccctga ttctgtggat aaccgtatta ccgcctttga gtgagctgat accgctcgcc 7320

gcagccgaac gaccgagcgc agcgagtcag tgagcgagga agcggaagag cgccagaagg 7380

ccgccagaga ggccgagcgc ggccgtgagg cttggacgct agggcagggc atgaaaaagc 7440

ccgtagcggg ctgctacggg cgtctgacgc ggtggaaagg gggaggggat gttgtctaca 7500

tggctctgct gtagtgagtg ggttgcgctc cggcagcggt cctgatcaat cgtcaccctt 7560

tctcggtcct tcaacgttcc tgacaacgag cctccttttc gccaatccat cgacaatcac 7620

cgcgagtccc tgctcgaacg ctgcgtccgg accggcttcg tcgaaggcgt ctatcgcggc 7680

ccgcaacagc ggcgagagcg gagcctgttc aacggtgccg ccgcgctcgc cggcatcgct 7740

gtcgccggcc tgctcctcaa gcacggcccc aacagtgaag tagctgattg tcatcagcgc 7800

attgacggcg tccccggccg aaaaacccgc ctcgcagagg aagcgaagct gcgcgtcggc 7860

cgtttccatc tgcggtgcgc ccggtcgcgt gccggcatgg atgcgcgcgc catcgcggta 7920

ggcgagcagc gcctgcctga agctgcgggc attcccgatc agaaatgagc gccagtcgtc 7980

gtcggctctc ggcaccgaat gcgtatgatt ctccgccagc atggcttcgg ccagtgcgtc 8040

gagcagcgcc cgcttgttcc tgaagtgcca gtaaagcgcc ggctgctgaa cccccaaccg 8100

ttccgccagt ttgcgtgtcg tcagaccgtc 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 acgctgtcgg catcggcctg attatgccgg tgctgccggg cctcctgcgc 8400

gatctggttc actcgaacga cgtcaccgcc cactatggca ttctgctggc gctgtatgcg 8460

ttggtgcaat ttgcctgcgc acctgtgctg ggcgcgctgt cggatcgttt cgggcggcgg 8520

ccaatcttgc tcgtctcgct ggccggcgcc actgtcgact acgccatcat ggcgacagcg 8580

-continued

```
cctttccttt gggttctcta tatcgggcgg atcgtggccg gcatcaccgg ggcgactggg   8640
gcggtagccg gcgcttatat tgccgatatc actgatggcg atgagcgcgc gcggcacttc   8700
ggcttcatga gcgcctgttt cgggttcggg atggtcgcgg gacctgtgct cggtgggctg   8760
atgggcggtt tctcccccca cgctccgttc ttcgccgcgg cagccttgaa cggcctcaat   8820
ttcctgacgg gctgtttcct tttgccggag tcgcacaaag gcgaacgccg gccgttacgc   8880
cgggaggctc tcaacccgct cgcttcgttc cggtgggccc ggggcatgac cgtcgtcgcc   8940
gccctgatgg cggtcttctt catcatgcaa cttgtcggac aggtgccggc cgcgctttgg   9000
gtcattttcg gcgaggatcg ctttcactgg gacgcgacca cgatcggcat ttcgcttgcc   9060
gcatttggca ttctgcattc actcgcccag gcaatgatca ccggccctgt agccgcccgg   9120
ctcggcgaaa ggcgggcact catgctcgga atgattgccg acggcacagg ctacatcctg   9180
cttgccttcg cgacacgggg atggatggcg ttcccgatca tggtcctgct tgcttcgggt   9240
ggcatcggaa tgccggcgct gcaagcaatg ttgtccaggc aggtggatga ggaacgtcag   9300
gggcagctgc aaggctcact ggcggcgctc accagcctga cctcgatcgt cggacccctc   9360
ctcttcacgg cgatctatgc ggcttctata acaacgtgga acgggtgggc atggattgca   9420
ggcgctgccc tctacttgct ctgcctgccg gcgctgcgtc gcgggctttg gagcggcgca   9480
gggcaacgag ccgatcgctg atcgtggaaa cgataggcct atgccatgcg ggtcaaggcg   9540
acttccggca agctatacgc gccctaggag tgcggttgga acgttggccc agccagatac   9600
tcccgatcac gagcaggacg ccgatgattt gaagcgcact cagcgtctga tccaagaaca   9660
accatcctag caacacggcg gtccccgggc tgagaaagcc cagtaaggaa acaactgtag   9720
gttcgagtcg cgagatcccc cggaaccaaa ggaagtaggt taaacccgct ccgatcaggc   9780
cgagccacgc caggccgaga acattggttc ctgtaggcat cgggattggc ggatcaaaca   9840
ctaaagctac tggaacgagc agaagtcctc cggccgccag ttgccaggcg gtaaaggtga   9900
gcagaggcac gggaggttgc cacttgcggg tcagcacggt tccgaacgcc atggaaaccg   9960
cccccgccag gcccgctgcg acgccgacag gatctagcgc tgcgtttggt gtcaacacca  10020
acagcgccac gcccgcagtt ccgcaaatag cccccaggac cgccatcaat cgtatcgggc  10080
tacctagcag agcggcagag atgaacacga ccatcagcgg ctgcacagcg cctaccgtcg  10140
ccgcgacccc gcccggcagg cggtagaccg aaataaacaa caagctccag aatagcgaaa  10200
tattaagtgc gccgaggatg aagatgcgca tccaccagat tcccgttgga atctgtcgga  10260
cgatcatcac gagcaataaa cccgccggca acgcccgcag cagcataccg gcgacccctc  10320
ggcctcgctg ttcgggctcc acgaaaacgc cggacagatg cgccttgtga gcgtccttgg  10380
ggccgtcctc ctgtttgaag accgacagcc caatgatctc gccgtcgatg taggcgccga  10440
atgccacggc atctcgcaac cgttcagcga acgcctccat gggctttttc tcctcgtgct  10500
cgtaaacgga cccgaacatc tctggagctt tcttcagggc cgacaatcgg atctcgcgga  10560
aatcctgcac gtcggccgct ccaagccgtc gaatctgagc cttaatcaca attgtcaatt  10620
ttaatcctct gtttatcggc agttcgtaga gcgcgccgtg cgtcccgagc gatactgagc  10680
gaagcaagtg cgtcgagcag tgcccgcttg ttcctgaaat gccagtaaag cgctggctgc  10740
tgaaccccca gccggaactg accccacaag gccctagcgt ttgcaatgca ccaggtcatc  10800
attgacccag gcgtgttcca ccaggccgct gcctcgcaac tcttcgcagg cttcgccgac  10860
```

-continued

ctgctcgcgc cacttcttca cgcgggtgga atccgatccg cacatgaggc ggaaggtttc 10920

cagcttgagc gggtacggct cccggtgcga gctgaaatag tcgaacatcc gtcgggccgt 10980

cggcgacagc ttgcggtact tctcccatat gaatttcgtg tagtggtcgc cagcaaacag 11040

cacgacgatt tcctcgtcga tcaggacctg gcaacgggac gttttcttgc cacggtccag 11100

gacgcggaag cggtgcagca gcgacaccga ttccaggtgc ccaacgcggt cggacgtgaa 11160

gcccatcgcc gtcgcctgta ggcgcgacag gcattcctcg gccttcgtgt aataccggcc 11220

attgatcgac cagcccaggt cctggcaaag ctcgtagaac gtgaaggtga tcggctcgcc 11280

gataggggtg cgcttcgcgt actccaacac ctgctgccac accagttcgt catcgtcggc 11340

ccgcagctcg acgccggtgt aggtgatctt cacgtccttg ttgacgtgga aaatgacctt 11400

gttttgcagc gcctcgcgcg ggattttctt gttgcgcgtg gtgaacaggg cagagcgggc 11460

cgtgtcgttt ggcatcgctc gcatcgtgtc cggccacggc gcaatatcga acaaggaaag 11520

ctgcatttcc ttgatctgct gcttcgtgtg tttcagcaac gcggcctgct tggcctcgct 11580

gacctgtttt gccaggtcct cgccggcggt ttttcgcttc ttggtcgtca tagttcctcg 11640

cgtgtcgatg gtcatcgact tcgccaaacc tgccgcctcc tgttcgagac gacgcgaacg 11700

ctccacggcg gccgatggcg cgggcagggc aggggggagcc agttgcacgc tgtcgcgctc 11760

gatcttggcc gtagcttgct ggaccatcga gccgacggac tggaaggttt cgcggggcgc 11820

acgcatgacg gtgcggcttg cgatggtttc ggcatcctcg gcggaaaacc ccgcgtcgat 11880

cagttcttgc ctgtatgcct tccggtcaaa cgtccgattc attcaccctc cttgcgggat 11940

tgccccgact cacgccgggg caatgtgccc ttattcctga tttgacccgc ctggtgcctt 12000

ggtgtccaga taatccacct tatcggcaat gaagtcggtc ccgtagaccg tctggccgtc 12060

cttctcgtac ttggtattcc gaatcttgcc ctgcacgaat accagcgacc ccttgcccaa 12120

atacttgccg tgggcctcgg cctgagagcc aaaacacttg atgcggaaga agtcggtgcg 12180

ctcctgcttg tcgccggtcg tggccgcgcc aacctttgcg atccgcaagc gcgcggtcgc 12240

catcttcacg ctggaacagt acgtcgaggc gggcatcatg acccgcgagc aatacgaggt 12300

cattaaaagc gccgtgattg atgatatagc ggcccggctg ctcctggttc tcgcgcaccg 12360

aaatgggtga cttcaccccg cgctctttga tcgtggcacc gatttccgcg atgctctccg 12420

gggaaaagcc ggggttgtcg gccgtccgcg gctgatgcgg atcttcgtcg atcaggtcca 12480

ggtccagctc gatagggccg gaaccgccct gagacgccgc aggagcgtcc aggaggctcg 12540

acaggtcgcc gatgctatcc aaccccaggc cggacggctg cgccgcgcct gcggcttcct 12600

gagcggccgc agcggtgttt ttcttggtgg tcttggcttg agccgcagtc attgggaaat 12660

ctccatcttc gtgaacacgt aatcagccag ggcgcgaacc tctttcgatg ccttgcgcgc 12720

ggccgttttc ttgatcttcc agaccggcac accggatgcg agggcatcgg cgatgctgct 12780

gcgcaggcca acggtggccg gaatcatcat cttggggtac gcggccagca gctcggcttg 12840

gtggcgcgcg tggcgcggat tccgcgcatc gaccttgctg ggcaccatgc caaggaattg 12900

cagcttggcg ttcttctggc gcacgttcgc aatggtcgtg accatcttct tgatgccctg 12960

gatgctgtac gcctcaagct cgatgggggа cagcacatag tcggccgcga agagggcggc 13020

cgccaggccg acgccaaggg tcggggccgt gtcgatcagg cacacgtcga agccttggtt 13080

cgccagggcc ttgatgttcg ccccgaacag ctcgcgggcg tcgtccagcg acagccgttc 13140

-continued

ggcgttcgcc agtaccgggt tggactcgat gagggcgagg cgcgcggcct ggccgtcgcc 13200

ggctgcgggt gcggtttcgg tccagccgcc ggcagggaca gcgccgaaca gcttgcttgc 13260

atgcaggccg gtagcaaagt ccttgagcgt gtaggacgca ttgccctggg ggtccaggtc 13320

gatcacggca acccgcaagc cgcgctcgaa aaagtcgaag gcaagatgca caagggtcga 13380

agtcttgccg acgccgcctt tctggttggc cgtgaccaaa gttttcatcg tttggtttcc 13440

tgtttttttct tggcgtccgc ttcccacttc cggacgatgt acgcctgatg ttccggcaga 13500

accgccgtta cccgcgcgta cccctcgggc aagttcttgt cctcgaacgc ggcccacacg 13560

cgatgcaccg cttgcgacac tgcgcccctg gtcagtccca gcgacgttgc gaacgtcgcc 13620

tgtggcttcc catcgactaa gacgccccgc gctatctcga tggtctgctg ccccacttcc 13680

agcccctgga tcgcctcctg gaactggctt tcggtaagcc gtttcttcat ggataacacc 13740

cataatttgc tccgcgcctt ggttgaacat agcggtgaca gccgccagca catgagagaa 13800

gtttagctaa acatttctcg cacgtcaaca cctttagccg ctaaaactcg tccttggcgt 13860

aacaaaacaa aagcccggaa accgggcttt cgtctcttgc cgcttatggc tctgcacccg 13920

gctccatcac caacaggtcg cgcacgcgct tcactcggtt gcggatcgac actgccagcc 13980

caacaaagcc ggttgccgcc gccgccagga tcgcgccgat gatgccggcc acaccggcca 14040

tcgcccacca ggtcgccgcc ttccggttcc attcctgctg gtactgcttc gcaatgctgg 14100

acctcggctc accataggct gaccgctcga tggcgtatgc cgcttctccc cttggcgtaa 14160

aacccagcgc cgcaggcggc attgccatgc tgcccgccgc tttcccgacc acgacgcgcg 14220

caccaggctt gcggtccaga ccttcggcca cggcgagctg cgcaaggaca taatcagccg 14280

ccgacttggc tccacgcgcc tcgatcagct cttgcactcg 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggggg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 720

-continued

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg 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

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggcgactag gggaggatag taatgctcgt 2580

caagacgcac atcacgaaga ttggggtcac gctgttcgcc gtggctctgt tctacgggtt 2640

catctacatg ctttccaact cgctctttgc gacgcggccg gccaccgccg ttgcggtcgg 2700

tgcggacggg aaagccctga ctagttccat atgatagatc tcaccaccac caccaccacc 2760

actaataggc cggccctccg tcgcgggcgg cacccacgcc cgcatcgatt ccaaggttca 2820

gccattgaga cggctccgct tcgcgcgcaa gcgcgggttg ggccgactgc aagcggagag 2880

ggaagcatgg cactgctcag cttcgagcga aaatatcgcg tgccgggggg cacgctggtc 2940

ggcggaaacc tgttcgactt ctgggtcggc cctttctatg tcggcttctt cggggttgcg 3000

-continued

```
acgtttttct tcgcggccct gggtatcatt ctgattgcct ggagtgccgt actccagggt   3060

acctggaacc cccaactcat ctctgtctac ccgccggccc ttgaatatgg cctgggaggt   3120

gcacccctcg caaaaggcgg gctgtggcag atcatcacga tctgcgccac tggtgccttc   3180

gtcagctggg cgctgcgcga agtcgaaatc tgccgtaagc tgggcatcgg gtaccacatc   3240

ccgttcgcct tcgcgttcgc catcctggcc tacctgacgc tggtgctgtt ccgcccggtg   3300

atgatgggcg cctggggcta tgccttcccc tacgggatct ggacgcacct cgactgggtg   3360

tcgaacacgg gctacaccta cggcaacttc cactacaacc ctgcccacat gatcgccatc   3420

tcgttcttct tcacgaacgc gctggctctg gcgctgcacg gcgcccttgt gctctccgcg   3480

gccaaccccg agaagggcaa ggaaatgcgg acgccggatc acgaggatac gttcttccgc   3540

gatctggtcg gctactcgat cgggacgctc ggcatccacc gcctcggcct gctgctctcg   3600

ctgagcgccg tcttcttcag cgccctctgc atgatcatta ccggcaccat ctggttcgat   3660

cagtgggtcg actggtggca atggtgggtg aagctgccgt ggtgggcgaa catcccggga   3720

ggcatcaatg gctgagtatc agaacatctt ctcccaggtc caggtccgcg gaccggccga   3780

cctggggatg accgaagacg tcaacctggc caaccgttcg ggcgtcggtc ccttctcgac   3840

cctgctcggc tggttcggca acgcccagct cggcccgatc tatctcggct cgctcggcgt   3900

cctgtccctc ttctcgggcc tgatgtggtt cttcaccatc gggatctggt tctggtatca   3960

ggcgggctgg aacccggccg tcttcctgcg cgacctgttc ttcttctcgc tcgagccgcc   4020

ggcacccgaa tacggtctgt ccttcgcggc tccgctgaag gaaggcgggc tgtggctgat   4080

cgcgtcgttc ttcatgttcg tcgcggtctg gtcctggtgg ggccgcacct atctccgcgc   4140

tcaggcgctg ggcatgggca agcacaccgc ctgggcgttc ctctcggcca tctggctgtg   4200

gatggtgctg ggcttcatcc gtccgatcct catggggtcc tggtcggaag cggttcccta   4260

cggcatcttc tcgcacctcg actggacgaa caacttctcg ctcgtccacg gcaacctgtt   4320

ctacaacccc ttccacggtc tctcgatcgc cttcctctac gggtcggccc tgctcttcgc   4380

gatgcacggt gcgaccatcc tcgcggtctc ccgcttcggc ggcgagcgcg agctggagca   4440

gatcgccgac cgcgggacgg cagcggagcg ggccgccctc ttctggcgct ggaccatggg   4500

tttcaacgcc acgatggaag gcatccaccg ctgggccatc tggatggcgg tcctcgtgac   4560

cctcaccggc ggcatcggga tcctgctctc gggcacggtc gtggacaact ggtacgtctg   4620

gggccagaac cacggcatgg cgccgctgaa ctgaggagcg atcacaatgg ctgacaagac   4680

catcttcaac gatcacctca acaccaatcc gaagaccaac cttcgcctct gggtcgcttt   4740

ccagatgatg aagggtgcgg gctgggctgg cggcgtgttc ttcgggacgc tccttctcat   4800

cgggttcttc cgggtggtcg ggcggatgct tccgatccag gagaaccagg ctccggcgcc   4860

gaacatcacc ggcgctctgg agaccgggat cgagctgatc aagcatctcg tctgagacaa   4920

gtctcggggc agggcggcgc gaggccgccc gctcctccaa gtccgggcca tatcgccggc   4980

ccgggtccgg ggcgacacca cagcccggtt cccttcctgt tggcgacagg gacctggtgc   5040

cgtgtggaag accgcacggc acccttttga cattcacggg aggctctgat gaccaatccc   5100

accccgcgac ccgaaacccc gcttttggat cgcgtctgct gcccggccga catgaaggcg   5160

ctgagtgacg ccgaactgga gcggctggcc gacgaagtgc gttccgaggt cagtgatagg   5220

ggtagtttct tattttaggc agtttatatg aaattaagac atgcagatgt cacagtggat   5280
```

-continued

attgaactgg tctcgaaagc tcaatatccc ccaaagcaca agcacaaact tcgacatcat 5340

gcagaagcgt ttcccgaacc gcgtcttcga cgtgggcatc gccgagcagc atgccgtgac 5400

cttcgcggcc ggcctcgccg gggccgggat gaagcccttc tgcgcgatct attcctcgtt 5460

cctgcaacgg ggttacgacc agatcgccca tgacgtggcg ctgcagaacc ttcccgtccg 5520

cttcgtgatc gaccgggcgg ggctcgtggg ggccgatggc gcgacccatg cgggggcctt 5580

cgacgttggc ttcatcactt cgctgcccaa catgaccgtg atggccgcgg ccgacgaggc 5640

cgagctcatc cacatgatcg ccaccgccgt ggccttcgac gagggcccca tcgccttccg 5700

cttcccgcgg ggcgaggggg tgggcgtcga gatgcccgag cgcgggacgg tgctggagcc 5760

cggccggggc cgcgtggtgc gcgaagggac ggatgtcgcg atcctctcct tcggcgcgca 5820

tctgcacgag gccttgcagg cggcgaaact tctcgaggcc gaggggggtga gcgtgaccgt 5880

ggccgacgcc cgcttctcgc gcccgctcga cacggggctc atcgaccagc tcgtgcgcca 5940

tcacgcggcg ctggtaacgg tggagcaggg ggccatgggc ggcttcggcg cccatgtcat 6000

gcactatctc gccaattccg gcggcttcga cgggggcctc gcgctccggg tcatgacgct 6060

gcccgaccgc ttcatcgagc aggcgagccc cgaggacatg tatgccgatg cggggctgcg 6120

ggccgaggat atcaagcttg gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt 6180

atccgctcac aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctggggtg 6240

cctaatgagt gagctaactc acattaattg cgttgcgctc actgcccgct ttccagtcgg 6300

gaaacctgtc gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc 6360

gtattgggcg ctcggtcttg ccttgctcgt cggtgatgta cttcaccagc tccgcgaagt 6420

cgctcttctt gatggagcgc atggggacgt gcttggcaat cacgcgcacc ccccggccgt 6480

tttagcggct aaaaaagtca tggctctgcc ctcgggcgga ccacgcccat catgaccttg 6540

ccaagctcgt cctgcttctc ttcgatcttc gccagcaggg cgaggatcgt ggcatcaccg 6600

aaccgcgccg tgcgcgggtc gtcggtgagc cagagtttca gcaggccgcc caggcggccc 6660

aggtcgccat tgatgcgggc cagctcgcgg acgtgctcat agtccacgac gcccgtgatt 6720

ttgtagccct ggccgacggc cagcaggtag gccgacaggc tcatgccggc cgccgccgcc 6780

ttttcctcaa tcgctcttcg ttcgtctgga aggcagtaca ccttgatagg tgggctgccc 6840

ttcctggttg gcttggtttc atcagccatc cgcttgccct catctgttac gccggcggta 6900

gccggccagc ctcgcagagc aggattcccg ttgagcaccg ccaggtgcga ataagggaca 6960

gtgaagaagg aacacccgct cgcgggtggg cctacttcac ctatcctgcc cggctgacgc 7020

cgttggatac accaaggaaa gtctacacga acccttggc aaaatcctgt atatcgtgcg 7080

aaaaaggatg gatataccga aaaaatcgct ataatgaccc cgaagcaggg ttatgcagcg 7140

gaaaagcgcc acgcttcccg aagggagaaa ggcggacagg tatccggtaa gcggcagggt 7200

cggaacagga gagcgcacga gggagcttcc agggggaaac gcctggtatc tttatagtcc 7260

tgtcgggttt cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg 7320

gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct tttgctggcc 7380

ttttgctcac atgttctttc ctgcgttatc ccctgattct gtggataacc gtattaccgc 7440

ctttgagtga gctgataccg ctcgccgcag ccgaacgacc gagcgcagcg agtcagtgag 7500

cgaggaagcg gaagagcgcc agaaggccgc cagagaggcc gagcgcggcc gtgaggcttg 7560

-continued

gacgctaggg cagggcatga aaaagcccgt agcgggctgc tacgggcgtc tgacgcggtg 7620

gaaaggggga ggggatgttg tctacatggc tctgctgtag tgagtgggtt gcgctccggc 7680

agcggtcctg atcaatcgtc acccttttctc ggtccttcaa cgttcctgac aacgagcctc 7740

cttttcgcca atccatcgac aatcaccgcg agtccctgct cgaacgctgc gtccggaccg 7800

gcttcgtcga aggcgtctat cgcggcccgc aacagcggcg agagcggagc ctgttcaacg 7860

gtgccgccgc gctcgccggc atcgctgtcg ccggcctgct cctcaagcac ggccccaaca 7920

gtgaagtagc tgattgtcat cagcgcattg acggcgtccc cggccgaaaa acccgcctcg 7980

cagaggaagc gaagctgcgc gtcggccgtt tccatctgcg gtgcgcccgg tcgcgtgccg 8040

gcatggatgc gcgcgccatc gcggtaggcg agcagcgcct gcctgaagct gcgggcattc 8100

ccgatcagaa atgagcgcca gtcgtcgtcg gctctcggca ccgaatgcgt atgattctcc 8160

gccagcatgg cttcggccag tgcgtcgagc agcgcccgct tgttcctgaa gtgccagtaa 8220

agcgccggct gctgaacccc caaccgttcc gccagtttgc gtgtcgtcag accgtctacg 8280

ccgacctcgt tcaacaggtc cagggcggca cggatcactg tattcggctg caactttgtc 8340

atgcttgaca ctttatcact gataaacata atatgtccac caacttatca gtgataaaga 8400

atccgcgcgt tcaatcggac cagcggaggc tggtccggag gccagacatg aaacccaaca 8460

tacccctgat cgtaattctg agcactgtcg cgctcgacgc tgtcggcatc ggcctgatta 8520

tgccggtgct gccgggcctc ctgcgcgatc tggttcactc gaacgacgtc accgcccact 8580

atggcattct gctggcgctg tatgcgttgg tgcaatttgc ctgcgcacct gtgctgggcg 8640

cgctgtcgga tcgtttcggg cggcggccaa tcttgctcgt ctcgctggcc ggcgccactg 8700

tcgactacgc catcatggcg acagcgcctt tcctttgggt tctctatatc gggcggatcg 8760

tggccggcat caccggggcg actggggcgg tagccggcgc ttatattgcc gatatcactg 8820

atggcgatga gcgcgcgcgg cacttcggct tcatgagcgc ctgtttcggg ttcgggatgg 8880

tcgcgggacc tgtgctcggt gggctgatgg gcggtttctc cccccacgct ccgttcttcg 8940

ccgcggcagc cttgaacggc ctcaatttcc tgacgggctg tttccttttg ccggagtcgc 9000

acaaaggcga acgccggccg ttacgccggg aggctctcaa cccgctcgct tcgttccggt 9060

gggcccgggg catgaccgtc gtcgccgccc tgatggcggt cttcttcatc atgcaacttg 9120

tcggacaggt gccggccgcg ctttgggtca ttttcggcga ggatcgcttt cactgggacg 9180

cgaccacgat cggcatttcg cttgccgcat ttggcattct gcattcactc gcccaggcaa 9240

tgatcaccgg ccctgtagcc gcccggctcg gcgaaaggcg ggcactcatg ctcggaatga 9300

ttgccgacgg cacaggctac atcctgcttg ccttcgcgac acggggatgg atggcgttcc 9360

cgatcatggt cctgcttgct tcgggtggca tcggaatgcc ggcgctgcaa gcaatgttgt 9420

ccaggcaggt ggatgaggaa cgtcaggggc agctgcaagg ctcactggcg gcgctcacca 9480

gcctgacctc gatcgtcgga cccctcctct tcacggcgat ctatgcggct tctataacaa 9540

cgtggaacgg gtgggcatgg attgcaggcg ctgccctcta cttgctctgc ctgccggcgc 9600

tgcgtcgcgg gctttggagc ggcgcagggc aacgagccga tcgctgatcg tggaaacgat 9660

aggcctatgc catgcgggtc aaggcgactt ccggcaagct atacgcgccc taggagtgcg 9720

gttggaacgt tggcccagcc agatactccc gatcacgagc aggacgccga tgatttgaag 9780

cgcactcagc gtctgatcca agaacaacca tcctagcaac acggcggtcc 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 aacgccatgg aaaccgcccc cgccaggccc gctgcgacgc cgacaggatc 10140

tagcgctgcg tttggtgtca acaccaacag cgccacgccc gcagttccgc aaatagcccc 10200

caggaccgcc atcaatcgta tcgggctacc tagcagagcg gcagagatga acacgaccat 10260

cagcggctgc acagcgccta ccgtcgccgc gaccccgccc ggcaggcggt agaccgaaat 10320

aaacaacaag ctccagaata gcgaaatatt aagtgcgccg aggatgaaga tgcgcatcca 10380

ccagattccc gttggaatct gtcggacgat catcacgagc aataaacccg ccggcaacgc 10440

ccgcagcagc ataccggcga cccctcggcc tcgctgttcg ggctccacga aaacgccgga 10500

cagatgcgcc ttgtgagcgt ccttggggcc gtcctcctgt ttgaagaccg acagcccaat 10560

gatctcgccg tcgatgtagg cgccgaatgc cacggcatct cgcaaccgtt cagcgaacgc 10620

ctccatgggc tttttctcct cgtgctcgta aacggacccg aacatctctg gagctttctt 10680

cagggccgac aatcggatct cgcggaaatc ctgcacgtcg gccgctccaa gccgtcgaat 10740

ctgagcctta atcacaattg tcaattttaa tcctctgttt atcggcagtt cgtagagcgc 10800

gccgtgcgtc ccgagcgata ctgagcgaag caagtgcgtc gagcagtgcc cgcttgttcc 10860

tgaaatgcca gtaaagcgct ggctgctgaa cccccagccg gaactgaccc cacaaggccc 10920

tagcgtttgc aatgcaccag gtcatcattg acccaggcgt gttccaccag gccgctgcct 10980

cgcaactctt cgcaggcttc gccgacctgc tcgcgccact tcttcacgcg ggtggaatcc 11040

gatccgcaca tgaggcggaa ggtttccagc ttgagcgggt acggctcccg gtgcgagctg 11100

aaatagtcga acatccgtcg ggccgtcggc gacagcttgc ggtacttctc ccatatgaat 11160

ttcgtgtagt ggtcgccagc aaacagcacg acgatttcct cgtcgatcag gacctggcaa 11220

cgggacgttt tcttgccacg gtccaggacg cggaagcggt gcagcagcga caccgattcc 11280

aggtgcccaa cgcggtcgga cgtgaagccc atcgccgtcg cctgtaggcg cgacaggcat 11340

tcctcggcct tcgtgtaata ccggccattg atcgaccagc ccaggtcctg gcaaagctcg 11400

tagaacgtga aggtgatcgg ctcgccgata ggggtgcgct tcgcgtactc caacacctgc 11460

tgccacacca gttcgtcatc gtcggcccgc agctcgacgc cggtgtaggt gatcttcacg 11520

tccttgttga cgtggaaaat gaccttgttt tgcagcgcct cgcgcgggat tttcttgttg 11580

cgcgtggtga acagggcaga gcgggccgtg tcgtttggca tcgctcgcat cgtgtccggc 11640

cacggcgcaa tatcgaacaa ggaaagctgc atttccttga tctgctgctt cgtgtgtttc 11700

agcaacgcgg cctgcttggc ctcgctgacc tgttttgcca ggtcctcgcc ggcggttttt 11760

cgcttcttgg tcgtcatagt tcctcgcgtg tcgatggtca tcgacttcgc caaacctgcc 11820

gcctcctgtt cgagacgacg cgaacgctcc acggcggccg atggcgcggg cagggcaggg 11880

ggagccagtt gcacgctgtc gcgctcgatc ttggccgtag cttgctggac catcgagccg 11940

acggactgga aggtttcgcg gggcgcacgc atgacggtgc ggcttgcgat ggtttcggca 12000

tcctcggcgg aaaaccccgc gtcgatcagt tcttgcctgt atgccttccg gtcaaacgtc 12060

cgattcattc accctccttg cgggattgcc ccgactcacg ccggggcaat gtgcccttat 12120

-continued

tcctgatttg acccgcctgg tgccttggtg tccagataat ccaccttatc ggcaatgaag 12180

tcggtcccgt agaccgtctg gccgtccttc tcgtacttgg tattccgaat cttgccctgc 12240

acgaatacca gcgacccctt gcccaaatac ttgccgtggg cctcggcctg agagccaaaa 12300

cacttgatgc ggaagaagtc ggtgcgctcc tgcttgtcgc cggtcgtggc cgcgccaacc 12360

tttgcgatcc gcaagcgcgc ggtcgccatc ttcacgctgg aacagtacgt cgaggcgggc 12420

atcatgaccc gcgagcaata cgaggtcatt aaaagcgccg tgattgatga tatagcggcc 12480

cggctgctcc tggttctcgc gcaccgaaat gggtgacttc accccgcgct ctttgatcgt 12540

ggcaccgatt tccgcgatgc tctccgggga aaagccgggg ttgtcggccg tccgcggctg 12600

atgcggatct tcgtcgatca ggtccaggtc cagctcgata gggccggaac cgccctgaga 12660

cgccgcagga gcgtccagga ggctcgacag gtcgccgatg ctatccaacc ccaggccgga 12720

cggctgcgcc gcgcctgcgg cttcctgagc ggccgcagcg gtgtttttct tggtggtctt 12780

ggcttgagcc gcagtcattg ggaaatctcc atcttcgtga acacgtaatc agccagggcg 12840

cgaacctctt tcgatgcctt gcgcgcggcc gttttcttga tcttccagac cggcacaccg 12900

gatgcgaggg catcggcgat gctgctgcgc aggccaacgg tggccggaat catcatcttg 12960

gggtacgcgg ccagcagctc ggcttggtgg cgcgcgtggc gcggattccg cgcatcgacc 13020

ttgctgggca ccatgccaag gaattgcagc ttggcgttct tctggcgcac gttcgcaatg 13080

gtcgtgacca tcttcttgat gccctggatg ctgtacgcct caagctcgat gggggacagc 13140

acatagtcgg ccgcgaagag ggcggccgcc aggccgacgc caagggtcgg ggccgtgtcg 13200

atcaggcaca cgtcgaagcc ttggttcgcc agggccttga tgttcgcccc gaacagctcg 13260

cgggcgtcgt ccagcgacag ccgttcggcg ttcgccagta ccgggttgga ctcgatgagg 13320

gcgaggcgcg cggcctggcc gtcgccggct gcgggtgcgg tttcggtcca gccgccggca 13380

gggacagcgc cgaacagctt gcttgcatgc aggccggtag caaagtcctt gagcgtgtag 13440

gacgcattgc cctgggggtc caggtcgatc acggcaaccc gcaagccgcg ctcgaaaaag 13500

tcgaaggcaa gatgcacaag ggtcgaagtc ttgccgacgc cgcctttctg gttggccgtg 13560

accaaagttt tcatcgtttg gtttcctgtt ttttcttggc gtccgcttcc cacttccgga 13620

cgatgtacgc ctgatgttcc ggcagaaccg ccgttacccg cgcgtacccc tcgggcaagt 13680

tcttgtcctc gaacgcggcc cacacgcgat gcaccgcttg cgacactgcg cccctggtca 13740

gtcccagcga cgttgcgaac gtcgcctgtg gcttcccatc gactaagacg ccccgcgcta 13800

tctcgatggt ctgctgcccc acttccagcc cctggatcgc ctcctggaac tggctttcgg 13860

taagccgttt cttcatggat aacacccata atttgctccg cgccttggtt gaacatagcg 13920

gtgacagccg ccagcacatg agagaagttt agctaaacat ttctcgcacg tcaacacctt 13980

tagccgctaa aactcgtcct tggcgtaaca aaacaaaagc ccggaaaccg ggctttcgtc 14040

tcttgccgct tatggctctg cacccggctc catcaccaac aggtcgcgca cgcgcttcac 14100

tcggttgcgg atcgacactg ccagcccaac aaagccggtt gccgccgccg ccaggatcgc 14160

gccgatgatg ccggccacac cggccatcgc ccaccaggtc gccgccttcc ggttccattc 14220

ctgctggtac tgcttcgcaa tgctggacct cggctcacca taggctgacc gctcgatggc 14280

gtatgccgct tctcccccttg gcgtaaaacc cagcgccgca ggcggcattg ccatgctgcc 14340

cgccgctttc ccgaccacga cgcgcgcacc aggcttgcgg tccagacctt cggccacggc 14400

-continued

gagctgcgca aggacataat cagccgccga cttggctcca cgcgcctcga tcagctcttg 14460

cactcgcgcg aaatccttgg cctccacggc cgccatgaat cgcgcacgcg gcgaaggctc 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

ccacccaggc cgccgccctc actgcccggc acctggtcgc tgaatgtcga tgccagcacc 60

tgcggcacgt caatgcttcc gggcgtcgcg ctcgggctga tcgcccatcc cgttactgcc 120

ccgatcccgg caatggcaag gactgccagc gctgccattt ttggggtgag gccgttcgcg 180

gccgaggggc gcagcccctg gggggatggg aggcccgcgt tagcgggccg ggagggttcg 240

agaaggggcg gcacccccct tcggcgtgcg cggtcacgcg cacagggcgc agccctggtt 300

aaaaacaagg tttataaata ttggtttaaa agcaggttaa aagacaggtt agcggtggcc 360

gaaaaacggg cggaaaccct tgcaaatgct ggattttctg cctgtggaca gcccctcaaa 420

tgtcaatagg tgcgcccctc atctgtcagc actctgcccc tcaagtgtca aggatcgcgc 480

ccctcatctg tcagtagtcg cgcccctcaa gtgtcaatac cgcagggcac ttatccccag 540

gcttgtccac atcatctgtg ggaaactcgc gtaaaatcag gcgttttcgc cgatttgcga 600

ggctggccag ctccacgtcg ccggccgaaa tcgagcctgc ccctcatctg tcaacgccgc 660

gccgggtgag tcggcccctc aagtgtcaac gtccgcccct catctgtcag tgagggccaa 720

gttttccgcg aggtatccac aacgccggcg gccgcggtgt ctcgcacacg gcttcgacgg 780

cgtttctggc gcgtttgcag ggccatagac ggccgccagc ccagcggcga gggcaaccag 840

cccggtgagc gtcggaaagg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg 900

tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 960

aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 1020

gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg ccccccctgac gagcatcaca 1080

aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 1140

ttcccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 1200

tgtccgcctt tctcccttcg ggaagcgtgg cgccattcgc cattcaggct gcgcaactgt 1260

tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa aggggatgt 1320

gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg ttgtaaaacg 1380

acggccagtg aattcggccg cgggctggcc gaggtgctgg gcaagcccta cctccaggcc 1440

cccatcgggg tcgagagcac gaccgccttc ctgcgccgcc tgggcgagat tctgggcctc 1500

gatccggagc ccttcatcga gcgcgagaag cactcgacgc tgaagcccgt gtgggatctg 1560

tggcggagtg tcacgcagga cttcttcggg acggccaatt tcggaatcgt ggcgaccgaa 1620

acttatgcaa gaggcatccg aaactatctc gaaggcgatc tcgggctgcc ctgcgccttc 1680

gccgtggccc gcaagagggg ctcgaagacc gacaacgaag cggtgcgcgg actgatccgc 1740

-continued

cagcaccgtc cgctcgtgct catggggtcg atcaacgaga agatttacct tgcggaactg 1800

aaagccggtc acggcccgca accctctttc atcgctgcct ctttcccggg tgcggcgatc 1860

cggcgcgcta ccggaacgcc cgttatggga tatgcaggtg ctacgtggtt actgcaggaa 1920

gtttgcaacg ccctgttcga cgccctgttc cacattctgc ccctcgggac ggagatggac 1980

agcgccgccg ccacaccgac gacactgcgc cgcgacttcc cgtgggatgc cgatgcgcaa 2040

gcggccctgg accgcatcgt agaggagcat ccggttctca cccggatcag cgccgcgcgt 2100

gccttgcgcg acgccgccga gaaggctgcc ctcgatgccg gtgccgagag ggtcgtgaga 2160

gagactgtcg aagccctgcg tgggccgggc ttcggcgaga ggaagggaga gaaccaatga 2220

gcgatcatgc cgtcaacacg ccggtccatg ccgccagggc ccacgggcac cgagcaccac 2280

gtgccgagtt ctacgtctac ttcgccgtca ttctgctggg cgccttcccg gtggccttcg 2340

tgagctggat cgtctcgacg atccgccacc gcaggcttcc caagcgcggc cccttcgcgt 2400

ccgcctggtt cgatgccaag gcgatcacgc cgctgatttt ccgcgcctga ccgcaggtca 2460

ggttgcgaca cgccattcgt cgtctcccca aggggcggcg gattaatcgg gagggcatgg 2520

tgccttaccg taacccacgc caccagcatg tggcgactag gggaggatag taatgaagtt 2580

ccaagtcaag gccctcgccg ccatcgccgc attcgcggcg ctgccggcgc tcgcgcagac 2640

tagttccata tgatagatct caccaccacc accaccacca ctaataggcc ggccctccgt 2700

cgcgggcggc acccacgccc gcatcgattc caaggttcag ccattgagac ggctccgctt 2760

cgcgcgcaag cgcgggttgg gccgactgca agcggagagg gaagcatggc actgctcagc 2820

ttcgagcgaa aatatcgcgt gccgggggcc acgctggtcg gcggaaacct gttcgacttc 2880

tgggtcggcc ctttctatgt cggcttcttc ggggttgcga cgtttttctt cgcggccctg 2940

ggtatcattc tgattgcctg gagtgccgta ctccagggta cctggaaccc ccaactcatc 3000

tctgtctacc cgccggccct tgaatatggc ctgggaggtg cacccctcgc aaaaggcggg 3060

ctgtggcaga tcatcacgat 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 actacaaccc tgcccacatg atcgccatct cgttcttctt cacgaacgcg 3360

ctggctctgg cgctgcacgg cgcccttgtg ctctccgcgg ccaaccccga gaagggcaag 3420

gaaatgcgga cgccggatca cgaggatacg ttcttccgcg atctggtcgg ctactcgatc 3480

gggacgctcg gcatccaccg cctcggcctg ctgctctcgc tgagcgccgt cttcttcagc 3540

gccctctgca tgatcattac cggcaccatc tggttcgatc agtgggtcga ctggtggcaa 3600

tggtgggtga agctgccgtg gtgggcgaac atcccgggag gcatcaatgg ctgagtatca 3660

gaacatcttc tcccaggtcc aggtccgcgg accggccgac ctggggatga ccgaagacgt 3720

caacctggcc aaccgttcgg gcgtcggtcc cttctcgacc ctgctcggct ggttcggcaa 3780

cgcccagctc ggcccgatct atctcggctc gctcggcgtc ctgtccctct tctcgggcct 3840

gatgtggttc ttcaccatcg ggatctggtt ctggtatcag gcgggctgga acccggccgt 3900

cttcctgcgc gacctgttct tcttctcgct cgagccgccg gcacccgaat acggtctgtc 3960

cttcgcggct ccgctgaagg aaggcgggct gtggctgatc gcgtcgttct tcatgttcgt 4020

-continued

cgcggtctgg tcctggtggg gccgcaccta tctccgcgct caggcgctgg gcatgggcaa 4080

gcacaccgcc tgggcgttcc tctcggccat ctggctgtgg atggtgctgg gcttcatccg 4140

tccgatcctc atggggtcct ggtcggaagc ggttccctac ggcatcttct cgcacctcga 4200

ctggacgaac aacttctcgc tcgtccacgg caacctgttc tacaacccct tccacggtct 4260

ctcgatcgcc ttcctctacg ggtcggccct gctcttcgcg atgcacggtg cgaccatcct 4320

cgcggtctcc cgcttcggcg gcgagcgcga gctggagcag atcgccgacc gcgggacggc 4380

agcggagcgg gccgccctct tctggcgctg gaccatgggt ttcaacgcca cgatggaagg 4440

catccaccgc tgggccatct ggatggcggt cctcgtgacc ctcaccggcg gcatcgggat 4500

cctgctctcg ggcacggtcg tggacaactg gtacgtctgg ggccagaacc acggcatggc 4560

gccgctgaac tgaggagcga tcacaatggc tgacaagacc atcttcaacg atcacctcaa 4620

caccaatccg aagaccaacc ttcgcctctg ggtcgctttc cagatgatga agggtgcggg 4680

ctgggctggc ggcgtgttct tcgggacgct ccttctcatc gggttcttcc gggtggtcgg 4740

gcggatgctt ccgatccagg agaaccaggc tccggcgccg aacatcaccg gcgctctgga 4800

gaccgggatc gagctgatca agcatctcgt ctgagacaag tctcggggca gggcggcgcg 4860

aggccgcccg ctcctccaag tccgggccat atcgccggcc cgggtccggg gcgacaccac 4920

agcccggttc ccttcctgtt ggcgacaggg acctggtgcc gtgtggaaga ccgcacggca 4980

ccctttgac attcacggga ggctctgatg accaatccca ccccgcgacc cgaaaccccg 5040

cttttggatc gcgtctgctg cccggccgac atgaaggcgc tgagtgacgc cgaactggag 5100

cggctggccg acgaagtgcg ttccgaggtc agtgataggg gtagtttctt attttaggca 5160

gtttatatga aattaagaca tgcagatgtc acagtggata ttgaactggt ctcgaaagct 5220

caatatcccc caaagcacaa gcacaaactt cgacatcatg cagaagcgtt tcccgaaccg 5280

cgtcttcgac gtgggcatcg ccgagcagca tgccgtgacc ttcgcggccg gcctcgccgg 5340

ggccgggatg aagcccttct gcgcgatcta ttcctcgttc ctgcaacggg gttacgacca 5400

gatcgcccat gacgtggcgc tgcagaacct tcccgtccgc ttcgtgatcg accgggcggg 5460

gctcgtgggg gccgatggcg cgacccatgc gggggccttc gacgttggct tcatcacttc 5520

gctgcccaac atgaccgtga tggccgcggc cgacgaggcc gagctcatcc acatgatcgc 5580

caccgccgtg gccttcgacg agggccccat cgccttccgc ttcccgcggg gcgaggggggt 5640

gggcgtcgag atgcccgagc gcgggacggt gctggagccc ggccggggcc gcgtggtgcg 5700

cgaagggacg gatgtcgcga tcctctcctt cggcgcgcat ctgcacgagg ccttgcaggc 5760

ggcgaaactt ctcgaggccg agggggtgag cgtgaccgtg gccgacgccc gcttctcgcg 5820

cccgctcgac acggggctca tcgaccagct cgtgcgccat cacgcggcgc tggtaacggt 5880

ggagcagggg gccatgggcg gcttcggcgc ccatgtcatg cactatctcg ccaattccgg 5940

cggcttcgac gggggcctcg cgctccgggt catgacgctg cccgaccgct tcatcgagca 6000

ggcgagcccc gaggacatgt atgccgatgc ggggctgcgg gccgaggata tcaagcttgg 6060

cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca 6120

acatacgagc cggaagcata aagtgtaaag cctggggtgc ctaatgagtg agctaactca 6180

cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc 6240

attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggcgc tcggtcttgc 6300

-continued

cttgctcgtc ggtgatgtac ttcaccagct ccgcgaagtc gctcttcttg atggagcgca 6360

tggggacgtg cttggcaatc acgcgcaccc cccggccgtt ttagcggcta aaaaagtcat 6420

ggctctgccc tcgggcggac cacgcccatc atgaccttgc caagctcgtc ctgcttctct 6480

tcgatcttcg ccagcagggc gaggatcgtg gcatcaccga accgcgccgt gcgcgggtcg 6540

tcggtgagcc agagtttcag caggccgccc aggcggccca ggtcgccatt gatgcgggcc 6600

agctcgcgga cgtgctcata gtccacgacg cccgtgattt tgtagccctg gccgacggcc 6660

agcaggtagg ccgacaggct catgccggcc gccgccgcct tttcctcaat cgctcttcgt 6720

tcgtctggaa ggcagtacac cttgataggt gggctgccct tcctggttgg cttggtttca 6780

tcagccatcc gcttgccctc atctgttacg ccggcggtag ccggccagcc tcgcagagca 6840

ggattcccgt tgagcaccgc caggtgcgaa taagggacag tgaagaagga acacccgctc 6900

gcgggtgggc ctacttcacc tatcctgccc ggctgacgcc gttggataca ccaaggaaag 6960

tctacacgaa ccctttggca aaatcctgta tatcgtgcga aaaaggatgg atataccgaa 7020

aaaatcgcta taatgacccc gaagcagggt tatgcagcgg aaaagcgcca cgcttcccga 7080

agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag 7140

ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg 7200

acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga aaaacgccag 7260

caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc 7320

tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc 7380

tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgcca 7440

gaaggccgcc agagaggccg agcgcggccg tgaggcttgg acgctagggc agggcatgaa 7500

aaagcccgta gcgggctgct acgggcgtct gacgcggtgg aaaggggggag gggatgttgt 7560

ctacatggct ctgctgtagt gagtgggttg cgctccggca gcggtcctga tcaatcgtca 7620

ccctttctcg gtccttcaac gttcctgaca acgagcctcc ttttcgccaa tccatcgaca 7680

atcaccgcga gtccctgctc gaacgctgcg tccggaccgg cttcgtcgaa ggcgtctatc 7740

gcggcccgca acagcggcga gagcggagcc tgttcaacgg tgccgccgcg ctcgccggca 7800

tcgctgtcgc cggcctgctc ctcaagcacg gccccaacag tgaagtagct gattgtcatc 7860

agcgcattga cggcgtcccc ggccgaaaaa cccgcctcgc agaggaagcg aagctgcgcg 7920

tcggccgttt ccatctgcgg tgcgcccggt cgcgtgccgg catggatgcg cgcgccatcg 7980

cggtaggcga gcagcgcctg cctgaagctg cgggcattcc cgatcagaaa tgagcgccag 8040

tcgtcgtcgg ctctcggcac cgaatgcgta tgattctccg ccagcatggc ttcggccagt 8100

gcgtcgagca gcgcccgctt gttcctgaag tgccagtaaa gcgccggctg ctgaaccccc 8160

aaccgttccg ccagtttgcg tgtcgtcaga ccgtctacgc cgacctcgtt caacaggtcc 8220

agggcggcac ggatcactgt attcggctgc aactttgtca tgcttgacac tttatcactg 8280

ataaacataa tatgtccacc aacttatcag tgataaagaa tccgcgcgtt caatcggacc 8340

agcggaggct ggtccggagg ccagacatga aacccaacat acccctgatc gtaattctga 8400

gcactgtcgc gctcgacgct gtcggcatcg gcctgattat gccggtgctg ccgggcctcc 8460

tgcgcgatct ggttcactcg aacgacgtca ccgcccacta tggcattctg ctggcgctgt 8520

atgcgttggt gcaatttgcc tgcgcacctg tgctgggcgc gctgtcggat cgtttcgggc 8580

-continued

```
ggcggccaat cttgctcgtc tcgctggccg gcgccactgt cgactacgcc atcatggcga   8640

cagcgccttt cctttgggtt ctctatatcg ggcggatcgt ggccggcatc accggggcga   8700

ctggggcggt agccggcgct tatattgccg atatcactga tggcgatgag cgcgcgcggc   8760

acttcggctt catgagcgcc tgtttcgggt tcgggatggt cgcgggacct gtgctcggtg   8820

ggctgatggg cggtttctcc ccccacgctc cgttcttcgc cgcggcagcc ttgaacggcc   8880

tcaatttcct gacgggctgt ttccttttgc cggagtcgca caaaggcgaa cgccggccgt   8940

tacgccggga ggctctcaac ccgctcgctt cgttccggtg ggcccggggc atgaccgtcg   9000

tcgccgccct gatggcggtc ttcttcatca tgcaacttgt cggacaggtg ccggccgcgc   9060

tttgggtcat tttcggcgag gatcgctttc actgggacgc gaccacgatc ggcatttcgc   9120

ttgccgcatt tggcattctg cattcactcg cccaggcaat gatcaccggc cctgtagccg   9180

cccggctcgg cgaaaggcgg gcactcatgc tcggaatgat tgccgacggc acaggctaca   9240

tcctgcttgc cttcgcgaca cggggatgga tggcgttccc gatcatggtc ctgcttgctt   9300

cgggtggcat cggaatgccg gcgctgcaag caatgttgtc caggcaggtg gatgaggaac   9360

gtcaggggca gctgcaaggc tcactggcgg cgctcaccag cctgacctcg atcgtcggac   9420

ccctcctctt cacggcgatc tatgcggctt ctataacaac gtggaacggg tgggcatgga   9480

ttgcaggcgc tgccctctac ttgctctgcc tgccggcgct gcgtcgcggg ctttggagcg   9540

gcgcagggca acgagccgat cgctgatcgt ggaaacgata ggcctatgcc atgcgggtca   9600

aggcgacttc cggcaagcta tacgcgccct aggagtgcgg ttggaacgtt ggcccagcca   9660

gatactcccg atcacgagca ggacgccgat gatttgaagc gcactcagcg tctgatccaa   9720

gaacaaccat cctagcaaca cggcggtccc cgggctgaga aagcccagta aggaaacaac   9780

tgtaggttcg agtcgcgaga tcccccggaa ccaaaggaag taggttaaac ccgctccgat   9840

caggccgagc cacgccaggc cgagaacatt ggttcctgta ggcatcggga ttggcggatc   9900

aaacactaaa gctactggaa cgagcagaag tcctccggcc gccagttgcc aggcggtaaa   9960

ggtgagcaga ggcacgggag gttgccactt gcgggtcagc acggttccga acgccatgga  10020

aaccgccccc gccaggcccg ctgcgacgcc gacaggatct agcgctgcgt ttggtgtcaa  10080

caccaacagc gccacgcccg cagttccgca aatagccccc aggaccgcca tcaatcgtat  10140

cgggctacct agcagagcgg cagagatgaa cacgaccatc agcggctgca cagcgcctac  10200

cgtcgccgcg accccgcccg gcaggcggta gaccgaaata aacaacaagc tccagaatag  10260

cgaaatatta agtgcgccga ggatgaagat gcgcatccac cagattcccg ttggaatctg  10320

tcggacgatc atcacgagca ataaacccgc cggcaacgcc cgcagcagca taccggcgac  10380

ccctcggcct cgctgttcgg gctccacgaa aacgccggac agatgcgcct tgtgagcgtc  10440

cttggggccg tcctcctgtt tgaagaccga cagcccaatg atctcgccgt cgatgtaggc  10500

gccgaatgcc acggcatctc gcaaccgttc agcgaacgcc tccatgggct ttttctcctc  10560

gtgctcgtaa acggacccga acatctctgg agctttcttc agggccgaca atcggatctc  10620

gcggaaatcc tgcacgtcgg ccgctccaag ccgtcgaatc tgagccttaa tcacaattgt  10680

caattttaat cctctgttta tcggcagttc gtagagcgcg ccgtgcgtcc cgagcgatac  10740

tgagcgaagc aagtgcgtcg agcagtgccc gcttgttcct gaaatgccag taaagcgctg  10800

gctgctgaac ccccagccgg aactgacccc acaaggccct agcgtttgca atgcaccagg  10860
```

-continued

```
tcatcattga cccaggcgtg ttccaccagg ccgctgcctc gcaactcttc gcaggcttcg  10920
ccgacctgct cgcgccactt cttcacgcgg gtggaatccg atccgcacat gaggcggaag  10980
gtttccagct tgagcgggta cggctcccgg tgcgagctga aatagtcgaa catccgtcgg  11040
gccgtcggcg acagcttgcg gtacttctcc catatgaatt tcgtgtagtg gtcgccagca  11100
aacagcacga cgatttcctc gtcgatcagg acctggcaac gggacgtttt cttgccacgg  11160
tccaggacgc ggaagcggtg cagcagcgac accgattcca ggtgcccaac gcggtcggac  11220
gtgaagccca tcgccgtcgc ctgtaggcgc gacaggcatt cctcggcctt cgtgtaatac  11280
cggccattga tcgaccagcc caggtcctgg caaagctcgt agaacgtgaa ggtgatcggc  11340
tcgccgatag gggtgcgctt cgcgtactcc aacacctgct gccacaccag ttcgtcatcg  11400
tcggcccgca gctcgacgcc ggtgtaggtg atcttcacgt ccttgttgac gtggaaaatg  11460
accttgtttt gcagcgcctc gcgcgggatt ttcttgttgc gcgtggtgaa cagggcagag  11520
cgggccgtgt cgtttggcat cgctcgcatc gtgtccggcc acggcgcaat atcgaacaag  11580
gaaagctgca tttccttgat ctgctgcttc gtgtgtttca gcaacgcggc ctgcttggcc  11640
tcgctgacct gttttgccag gtcctcgccg gcggtttttc gcttcttggt cgtcatagtt  11700
cctcgcgtgt cgatggtcat cgacttcgcc aaacctgccg cctcctgttc gagacgacgc  11760
gaacgctcca cggcggccga tggcgcgggc gggcaggggg agccagttgc acgctgtcgc  11820
gctcgatctt ggccgtagct tgctggacca tcgagccgac ggactggaag gtttcgcggg  11880
gcgcacgcat gacggtgcgg cttgcgatgg tttcggcatc ctcggcggaa aaccccgcgt  11940
cgatcagttc ttgcctgtat gccttccggt caaacgtccg attcattcac cctccttgcg  12000
ggattgcccc gactcacgcc ggggcaatgt gcccttattc ctgatttgac ccgcctggtg  12060
ccttggtgtc cagataatcc accttatcgg caatgaagtc ggtcccgtag accgtctggc  12120
cgtccttctc gtacttggta ttccgaatct tgccctgcac gaataccagc gacccccttgc  12180
ccaaatactt gccgtgggcc tcggcctgag agccaaaaca cttgatgcgg aagaagtcgg  12240
tgcgctcctg cttgtcgccg gtcgtggccg cgccaacctt tgcgatccgc aagcgcgcgg  12300
tcgccatctt cacgctggaa cagtacgtcg aggcgggcat catgacccgc gagcaatacg  12360
aggtcattaa aagcgccgtg attgatgata tagcggcccg gctgctcctg gttctcgcgc  12420
accgaaatgg gtgacttcac cccgcgctct ttgatcgtgg caccgatttc cgcgatgctc  12480
tccggggaaa agccggggtt gtcggccgtc cgcggctgat gcggatcttc gtcgatcagg  12540
tccaggtcca gctcgatagg gccggaaccg ccctgagacg ccgcaggagc gtccaggagg  12600
ctcgacaggt cgccgatgct atccaacccc aggccggacg gctgcgccgc gcctgcggct  12660
tcctgagcgg ccgcagcggt gtttttcttg gtggtcttgg cttgagccgc agtcattggg  12720
aaatctccat cttcgtgaac acgtaatcag ccagggcgcg aacctctttc gatgccttgc  12780
gcgcggccgt tttcttgatc ttccagaccg gcacaccgga tgcgagggca tcggcgatgc  12840
tgctgcgcag gccaacggtg gccggaatca tcatcttggg gtacgcggcc agcagctcgg  12900
cttggtggcg cgcgtggcgc ggattccgcg catcgacctt gctgggcacc atgccaagga  12960
attgcagctt ggcgttcttc tggcgcacgt tcgcaatggt cgtgaccatc ttcttgatgc  13020
cctggatgct gtacgcctca agctcgatgg gggacagcac atagtcggcc gcgaagaggg  13080
cggccgccag gccgacgcca agggtcgggg ccgtgtcgat caggcacacg tcgaagcctt  13140
```

-continued

ggttcgccag ggccttgatg ttcgccccga acagctcgcg ggcgtcgtcc agcgacagcc 13200

gttcggcgtt cgccagtacc gggttggact cgatgagggc gaggcgcgcg gcctggccgt 13260

cgccggctgc gggtgcggtt tcggtccagc cgccggcagg gacagcgccg aacagcttgc 13320

ttgcatgcag gccggtagca aagtccttga gcgtgtagga cgcattgccc tgggggtcca 13380

ggtcgatcac ggcaacccgc aagccgcgct cgaaaaagtc gaaggcaaga tgcacaaggg 13440

tcgaagtctt gccgacgccg cctttctggt tggccgtgac caaagttttc atcgtttggt 13500

ttcctgtttt ttcttggcgt ccgcttccca cttccggacg atgtacgcct gatgttccgg 13560

cagaaccgcc gttacccgcg cgtacccctc gggcaagttc ttgtcctcga acgcggccca 13620

cacgcgatgc accgcttgcg acactgcgcc cctggtcagt cccagcgacg ttgcgaacgt 13680

cgcctgtggc ttcccatcga ctaagacgcc ccgcgctatc tcgatggtct gctgccccac 13740

ttccagcccc tggatcgcct cctggaactg gctttcggta agccgtttct tcatggataa 13800

cacccataat ttgctccgcg ccttggttga acatagcggt gacagccgcc agcacatgag 13860

agaagtttag ctaaacattt ctcgcacgtc aacaccttta gccgctaaaa ctcgtccttg 13920

gcgtaacaaa acaaaagccc ggaaaccggg ctttcgtctc ttgccgctta tggctctgca 13980

cccggctcca tcaccaacag gtcgcgcacg cgcttcactc ggttgcggat cgacactgcc 14040

agcccaacaa agccggttgc cgccgccgcc aggatcgcgc cgatgatgcc ggccacaccg 14100

gccatcgccc accaggtcgc cgccttccgg ttccattcct gctggtactg cttcgcaatg 14160

ctggacctcg gctcaccata ggctgaccgc tcgatggcgt atgccgcttc tcccccttggc 14220

gtaaaaccca gcgccgcagg cggcattgcc atgctgcccg ccgctttccc gaccacgacg 14280

cgcgcaccag gcttgcggtc cagaccttcg gccacggcga gctgcgcaag gacataatca 14340

gccgccgact tggctccacg cgcctcgatc agctcttgca ctcgcgcgaa atccttggcc 14400

tccacggccg 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 ttgtgcggcc 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 aacactatga agcgcaaagt 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 tgtgcggcct ccttactgta 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 aacactatg 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 ctcctgcct 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 aacactatg 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 cgacacttg 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
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 tggtggtggt ggtggtga 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

gtgttcctcc tgcctact 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 His His His His His His His His 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 His His His His His His His His His His 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.*

* * * * *