MINICELL BASED DELIVERY OF BIOLOGICALLY ACTIVE COMPOUNDS

Inventors: Roger A. Sabbadini, Lakeside, CA (US); Neil Berkley, San Diego, CA (US); Mark W. Surber, San Diego, CA (US)

Assignee: Vaxion Therapeutics, Inc., San Diego, CA (US)

Appl. No.: 13/294,920
Filed: Nov. 11, 2011

Related U.S. Application Data

Division of application No. 11/725,196, filed on Mar. 16, 2007, now Pat. No. 8,101,396, which is a division of application No. 11/580,095, filed on Oct. 11, 2006, now Pat. No. 7,871,815, which is a division of application No. 10/156,902, filed on May 28, 2002, now Pat. No. 7,183,105, which is a division of application No. 10/154,951, filed on May 24, 2002, now abandoned.

Provisional application No. 60/359,843, filed on Feb. 25, 2002, provisional application No. 60/293,566, filed on May 24, 2001.

Publication Classification

Int. Cl. C12N 1/20 (2006.01)
U.S. Cl. 435/252.1

ABSTRACT

Minicells are used to deliver biologically active compounds including radioisotopes, polypeptides, nucleic acids, small molecules, drug molecules, and chemotherapeutic agents. In some cases, the minicell displays ligands or binding moieties that target the minicell to a desired host cell.
Figure 1
<table>
<thead>
<tr>
<th>Rhamnose (mM)</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MalE(L)-NTR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MalE (host)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2**
MINICELL BASED DELIVERY OF BIOLOGICALLY ACTIVE COMPOUNDS

RELATED APPLICATIONS


FIELD OF THE INVENTION

[0002] The invention is drawn to compositions and methods for the production of a chromosomal archeabacterial, eubacterial and anucleate eukaryotic cells that are used as, e.g., therapeutics and/or diagnostics, reagents in drug discovery and functional proteomics, research tools, and in other applications as well.

BACKGROUND OF THE INVENTION

[0003] The following description of the background of the invention is provided to aid in understanding the invention, but is not admitted to describe or constitute prior art to the invention. The contents of the articles, patents, and patent applications, and all other documents and electronically available information mentioned or cited in this application, are hereby incorporated by reference in their entirety to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference. Applicants reserve the right to physically incorporate into this application any and all materials and information from any such articles, patents, patent applications, or other documents.


[0005] Prokaryotic (a.k.a. eubacterial) minicells have been used to produce various eubacterial proteins. See, e.g., Michael Gaide et al., The kdpF Subunit Is Part of the K+-translocating Kdp Complex of Escherichia coli and Is Responsible for Stabilization of the Complex in vitro, 274 (55) Jn. of Biological Chemistry 37901 (1999); Harlow et al., Cloning and Characterization of the gsk Gene Encoding Granuline Kinase of Escherichia coli, 177(8) J. of Bacteriology 2236 (1995); Carol L. Pickett et al., Cloning, Sequencing, and Expression of the Escherichia coli Cytolethal Distnding Toxin Genes, 62(3) Infection & Immunity 1046 (1994); Raimund Eck & Röm Belfer, Cloning and characterization of a gene coding for the catechol 1,2 dioxygenase of Arthrobacter sp. m3A3, 123 Gene 87 (1993); Andreas Schlösser et al., Subcloning, Nucleotide Sequence, and Expression of trkG, a Gene That Encodes an Integral Membrane Protein Involved in Potassium Uptake via the Trk System of Escherichia coli, 173(10) J. of Bacteriology 3170 (1991); Mehrdad Jannatipour et al., Translocation of Vibrio harveyi N’N’-Di-acetylchitobiose to the Outer Membrane of Escherichia coli 169(8) J. of Bacteriology 3785 (1987); and Jacobs et al., Expression of Mycobacterium leprae genes from a Streptococcus mutans promoter in Escherichia coli K-12, 83(6) Proc. Natl. Acad. Sci. USA 1926 (1986).

[0006] Various bacteria have been used, or proposed to be used, as gene delivery vectors to mammalian cells. For reviews, see Grillot-Courvalin et al., Bacteria as gene delivery vectors for mammalian cells, 10 Current Opinion in Biotechnology 477 (1999); Johnsen et al., Transfer of DNA from Genetically Modified Organisms (GMOs), Biotechnological Institute, 1-70 (2000); Sizemore et al., Attenuated Shigella as a DNA delivery vehicle for DNA-mediated immunization, 270(5234) Science 299 (1995); Patrice Courvalin et al., Gene transfer from bacteria to mammalian cells, 318 C. R. Acad. Sci. 1207 (1995); Sizemore et al. Attenuated bacteria as a DNA delivery vehicle for DNA-mediated immunization, 15(8) Vaccine 804 (1997).

[0007] U.S. Pat. No. 4,190,495, which issued Feb. 26, 1980, to Curtiss is drawn to minicell producing strains of E. coli that are stated to be useful for the recombinant expression of proteins.

[0008] U.S. Pat. No. 4,311,797, which issued Jan. 19, 1982 to Khachatourians is stated to be drawn to a minicell based vaccine. The vaccine is stated to induce the production of antibodies against enteropathogenic E. coli cells in cattle and is stated to be effective against coliform enteritis.

[0009] Eubacterial minicells expressing immunogens from other prokaryotes have been described. Purcell et al., Molecular cloning and characterization of the 15-kilodalton major immunogen of Treponema pallidum, Infect. Immun. 57:3708, 1989.

[0010] In “Biotechnology: Promise . . . and Peril” (IDRC Reports 9:4-7, 1980) authors Fleury and Shirkie aver that George Khachatourians at the University of Saskatchewan, Canada, “is working on a vaccine against cholera using ‘minicells.’” The minicells are said to contain “genes from the pathogenic agent,” and the pathogen antigens are carried on the surface of the minicells (p. 5, paragraph bridging the central and right columns).

[0011] Lundstrom et al., Secretion of Semliki Forest virus membrane glycoprotein E1 from Bacillus subtilis, Virus Res.
describe the expression of the E1 protein of the eukaryotic virus, Semliki Forest virus (SFV), in Bacillus minicells. The SFV E1 protein used in these studies is not the native E1 protein. Rather, it is a fusion protein in which the N-terminal signal sequence and C-terminal transmembrane domain have been removed and replaced with signal sequences from a gene from Bacillus amylophilipatriciens. The authors aver that “E1 is properly translocated through the cell membrane and secreted” (p. 81, 1.19-20), and note that “it has been difficult to express viral membrane proteins in prokaryotes” (p. 81, 1. 27).

U.S. Pat. No. 4,237,224, which issued Dec. 2, 1980, to Cohen and Boyer, describes the expression of X. laevis DNA in E. coli minicells.

U.S. patent application Ser. No. 60/293,566 (attorney docket Nos. 078853-0401 and 089608-0201), is entitled “Minicell Compositions and Methods,” and was filed May 24, 2001, by Sabbadin, Roger A., Berkley, Neil L., and Klepper, Robert E., and is hereby incorporated in its entirety by reference.


SUMMARY OF THE INVENTION

The invention is drawn to compositions and methods for the production and use of minicells, including but not limited to eubacterial minicells, in applications such as diagnostics, therapeutics, research, compound screening and drug discovery, as well as agents for the delivery of nucleic acids and other bioactive compounds to cells.

Minicells are derivatives of cells that lack chromosomal DNA and which are sometimes referred to as amucleate cells. Because eubacterial and archeabacterial cells, unlike eukaryotic cells, do not have a nucleus (a distinct organelle that contains chromosomes), these non-eukaryotic minicells are more accurately described as being “without chromosomes” or “achromosomal,” as opposed to “amucleate.” Nonetheless, those skilled in the art often use the term “amucleate” when referring to bacterial minicells in addition to other minicells. Accordingly, in the present disclosure, the term “minicells” encompasses derivatives of eubacterial cells that lack a chromosome; derivatives of archeabacterial cells that lack their chromosome(s), and amucleate derivatives of eukaryotic cells. It is understood, however, that some of the relevant art may use the terms “anucleate minicells” or “anucleate cells” loosely to refer to any of the preceding types of minicells.

In one aspect, the invention is drawn to a eubacterial minicell comprising a membrane protein that is not naturally found in a prokaryote, i.e., a membrane protein from an eukaryote or an archeabacterium. Such minicells may, but need not, comprise an expression element that encodes and expresses the membrane protein that it comprises. The membrane protein may be one found in any non-eubacterial membrane, including, by way of non-limiting example, a cellular membrane, a nuclear membrane, a nuclear envelope, a membrane of the endoplasmic reticulum (ER), a membrane of a Golgi body, a membrane of a lysosome a membrane of a peroxisome, a caveolar membrane, an outer membrane of a mitochondrion or a chloroplast, and an inner membrane of a mitochondrion or a chloroplast. By way of non-limiting example, a membrane protein may be a receptor, such as a G-protein coupled receptor; an enzyme, such as ATPase or adenylate cyclase, a cytochrome; a channel; a transporter; or a membrane-bound nucleic acid binding factor, such as a transcription and/or translation factor; signaling components; components of the electron transport chain (ETC); or cellular antigens. A membrane fusion protein, which is generated in vitro using molecular cloning techniques, does not occur in nature and is thus a membrane protein that is not naturally found in a prokaryote, even if the fusion protein is prepared using amino acid sequences derived from eubacterial proteins.

Minicells that have segregated from parent cells lack chromosomal and/or nuclear components, but retain the cytoplasm and its contents, including the cellular machinery required for protein expression. Although chromosomes do not segregate into minicells, extrachromosomal and/or episomal genetic expression elements will segregate, or may be introduced into minicells after segregation from parent cells. Thus, in one aspect, the invention is drawn to minicells comprising an expression element, which may be an inducible expression element, that comprises expression sequences operably linked to an open reading frame (ORF) that encodes the non-eubacterial membrane protein. In a related aspect, the invention is drawn to a method of making a eubacterial minicell comprising a membrane protein that is not naturally found in a prokaryote, the method comprising growing minicell-producing host cells, the host cells having an expression element, which may be an inducible expression element, that comprises expression sequences operably linked to an ORF that encodes a non-eubacterial membrane protein. In a related aspect, the invention is drawn to a method of making a eubacterial minicell comprising a membrane protein that is not naturally found in a prokaryote, the method comprising growing minicell-producing host cells, the host cells having an expression element, which may be an inducible expression element, that comprises expression sequences operably linked to an ORF that encodes a non-eubacterial membrane protein; and preparing minicells from the host cells. Optionally, at any point in the method, an inducing agent is provided in order to induce expression of an ORF that encodes a non-eubacterial membrane protein.

In one aspect, the invention is drawn to display produced membrane-associated protein(s) on the surface of the minicell. For purposes of this document, the term “display” is defined as exposure of the structure of interest on the outer surface of the minicell. By way of non-limiting example, this structure may be an internally expressed membrane protein or chimeric construct to be inserted in or associated with the minicell membrane such that the extracellular domain or domain of interest is exposed on the outer surface of the minicell (expressed and displayed on the surface of the minicell or expressed in the parental cell to be displayed on the surface of the segregated minicell). In any scenario, the “displayed” protein or protein domain is available for interaction with extracellular components. A membrane-associated protein may have more than one extracellular domain, and a minicell of the invention may display more than one membrane-associated protein.

A membrane protein displayed by eubacterial minicells may be a receptor. Receptors include, by way of non-limiting example, G-coupled protein receptors, hormone receptors, and growth factor receptors. Minicells displaying a receptor may, but need not, bind ligands of the receptor. In therapeutic applications of this aspect of the invention, the ligand is an undesirable compound that is bound to its receptor and, in some aspects, is internalized or inactivated by the
minicells. In drug discovery applications of this aspect of the invention, the ligand for the receptor may be detectably labeled so that its binding to its receptor may be quantified. In the latter circumstance, the minicells may be used to identify and isolate, from a pool of compounds, one or more compounds that inhibit or stimulate the activity of the receptor. That is, these minicells can be used in screening assays, including assays such as those used in high throughput screening (HTS) systems and other drug discovery methods, for the purpose of preparing compounds that influence the activity of a receptor of interest.

[0020] The displayed domain of a membrane protein may be an enzymatic domain such as having oxidoreductase, transferase, hydrolase, lyase, isomerase ligase, lipase, kinase, phosphatase, protease, nuclelease and/or synthetase activity. Contacting such minicells with the appropriate substrate of the enzyme allows the substrate to be converted to reactant. When either the substrate or reactant is detectable, the reaction catalyzed by the membrane-bound enzyme may be quantified. In the latter instance, the minicells may be used to identify and isolate, from a pool of compounds, one or more compounds that inhibit or stimulate the activity of the enzyme represented by the displayed enzymatic moiety. That is, these minicells can be used in screening assays, including assays such as those used in high throughput screening (HTS) systems and other drug discovery methods, for the purpose of preparing compounds that influence the activity of an enzyme or enzymatic moiety of interest.

[0021] The membrane protein displayed by minicells may be a fusion protein, i.e., a protein that comprises a first polypeptide having a first amino acid sequence and a second polypeptide having a second amino acid sequence, wherein the first and second amino acid sequences are not naturally present in the same polypeptide. At least one polypeptide in a membrane fusion protein is a "transmembrane domain" or "membrane-anchoring domain". The transmembrane and membrane-anchoring domains of a membrane fusion protein may be selected from membrane proteins that naturally occur in a eucaryote, such as a fungus, a unicellular eucaryote, a plant and an animal, such as a mammal including a human. Such domains may be from a viral membrane protein naturally found in a virus such as a bacteriophage or a eucaryotic virus, e.g., an adenovirus or a retrovirus. Such domains may be from a membrane protein naturally found in an archaeabacterium such as a thermophile.

[0022] The displayed domain of a membrane fusion protein may be an enzymatic domain such as having oxidoreductase, transferase, hydrolase, lyase, isomerase ligase, lipase, kinase, phosphatase, protease, nuclelease and/or synthetase activity. Contacting such minicells with the appropriate substrate of the enzyme allows the substrate to be converted to reactant. When either the substrate or reactant is detectable, the reaction catalyzed by the membrane-bound enzyme may be quantified. In the latter instance, the minicells may be used to identify and isolate, from a pool of compounds, one or more compounds that inhibit or stimulate the activity of the enzyme represented by the displayed enzymatic moiety. That is, these minicells can be used in screening assays, including assays such as those used in high throughput screening (HTS) systems and other drug discovery methods, for the purpose of preparing compounds that influence the activity of an enzyme or enzymatic moiety of interest.

[0023] The displayed domain of a membrane fusion protein may be a binding moiety. By way of non-limiting example, binding moieties used for particular purposes may be a binding moiety directed to a compound or moiety displayed by a specific cell type or cells found predominantly in one type of tissue, which may be used to target minicells and their contents to specific cell types or tissues; or a binding moiety that is directed to a compound or moiety displayed by a pathogen, which may be used in diagnostic or therapeutic methods; a binding moiety that is directed to an undesirable compound, such as a toxin, which may be used to bind and preferably internalize and/or neutralize the undesirable compound; a diseased cell; or the binding moiety may be a domain that allows for the minicells to be covalently or non-covalently attached to a support material, which may be used in compositions and methods for compound screening and drug discovery. By "diseased cell" it is meant pathogen-infected cells, malfunctioning cells, and dysfunctional cells, e.g., cancer cells.

[0024] In various aspects, the minicells of the invention comprise one or more biologically active compounds. The term "biologically active" (synonymous with "bioactive") indicates that a composition or compound itself has a biological effect, or that it modifies, causes, promotes, enhances, blocks, reduces, limits the production or activity of, or reacts with or binds to an endogenous molecule that has a biological effect. A "biological effect" may be but is not limited to one that stimulates or causes an immunoreactive response; one that impacts a biological process in an animal; one that impacts a biological process in a pathogen or parasite; one that generates or causes to be generated a detectable signal; and the like. Biologically active compositions, complexes or compounds may be used in therapeutic, prophylactic and diagnostic methods and compositions. Biologically active compositions, complexes or compounds set to cause or stimulate a desired effect upon an animal. Non-limiting examples of desired effects include, for example, preventing, treating or curing a disease or condition in an animal suffering therefrom; limiting the growth of or killing a pathogen in an animal infected thereby; augmenting the phenotype or genotype of an animal; stimulating a prophylactic immunoreactive response in an animal; or diagnosing a disease or disorder in an animal.

[0025] In the context of therapeutic applications of the invention, the term "biologically active" indicates that the composition, complex or compound has an activity that impacts an animal suffering from a disease or disorder in a positive sense and/or impacts a pathogen or parasite in a negative sense. Thus, a biologically active composition, complex or compound may cause or promote a biological or biochemical activity within an animal that is detrimental to the growth and/or maintenance of a pathogen or parasite; or of cells, tissues or organs of an animal that have abnormal growth or biochemical characteristics, such as cancer cells.

[0026] In the context of diagnostic applications of the invention, the term "biologically active" indicates that the composition, complex or compound can be used for in vivo or ex vivo diagnostic methods and in diagnostic compositions and kits. For diagnostic purposes, a preferred biologically active composition or compound is one that can be detected, typically (but not necessarily) by virtue of comprising a detectable polypeptide. Antibodies to an epitope found on composition or compound may also be used for its detection.

[0027] In the context of prophylactic applications of the invention, the term "biologically active" indicates that the composition or compound induces or stimulates an immu-
noreactive response. In some preferred embodiments, the immunoreactive response is designed to be prophylactic, i.e., prevents infection by a pathogen. In other preferred embodiments, the immunoreactive response is designed to cause the immune system of an animal to react to the detriment of cells of an animal, such as cancer cells, that have abnormal growth or biochemical characteristics. In this application of the invention, compositions, complexes or compounds comprising antigens are formulated as a vaccine.

[0028] It will be understood by those skilled in the art that a given composition, complex or compound may be biologically active in therapeutic, diagnostic and prophylactic applications. A composition, complex or compound that is described as being “biologically active in a cell” is one that has biological activity in vitro (i.e., in a cell culture) or in vivo (i.e., in the cells of an animal). A “biologically active component” of a composition or compound is a portion thereof that is biologically active once it is liberated from the composition or compound. It should be noted, however, that such a component may also be biologically active in the context of the composition or compound.

[0029] In one aspect, the minicells of the invention comprise a therapeutic agent. Such minicells may be used to deliver therapeutic agents. In a preferred embodiment, a minicell comprising a therapeutic agent displays a binding moiety that specifically binds a ligand present on the surface of a cell, so that the minicells may be “targeted” to the cell. The therapeutic agent may be any type of compound or moiety, including without limitation small molecules, polypeptides, antibodies and antibody derivatives and nucleic acids. The therapeutic agent may be a drug: a produrg, i.e., a compound that becomes biologically active in vivo after being introduced into a subject in need of treatment; or an immunogen.

[0030] In one aspect, the minicells of the invention comprise a detectable compound or moiety. As is understood by those of skill in the art, a compound or moiety that is “detectable” produces a signal that can be detected by spectroscopic, photochemical, biochemical, immunochemical, electromagnetic, radiochemical, or chemical means such as fluorescence, chemiluminescence, or chemiluminescence, electrochemical luminescence, or any other appropriate means. A detectable compound may be a detectable polypeptide, and such polypeptides may, but need not, be incorporated into fusion membrane proteins of the minicell. Detectable polypeptides or amino acid sequences, includes, by way of non-limiting example, a green fluorescent protein (GFP), a luciferase, a beta-galactosidase, a His tag, an epitope, or a biotin-binding protein such as streptavidin or avidin. The detectable compound or moiety may be a radiolabeled compound or radioisotope. A detectable compound or moiety may be a small molecule such as, by way of non-limiting example, a fluorescent dye; a radioactive isotope; or a compound that may be detected by x-rays or electromagnetic radiation. Image enhancers as those used for CAT and PET scans (e.g., calcium, gallium) may be used. In another non-limiting example, detectable labels may also include loss of catalytic substrate or gain of catalytic product following catalysis by a minicell displayed, solute cytoplasmic, or secreted enzyme.

[0031] In one aspect, the invention is drawn to a minicell comprising one or more bioactive nucleic acids or templates thereof. By way of non-limiting example, a bioactive nucleic acid may be an antisense oligonucleotide, an aptamer, an antisense transcript, a ribosomal RNA (rRNA), a transfer RNA (tRNA), a molecular decoy, or an enzymatically active nucleic acid, such as a ribozyme. Such minicells can, but need not, comprise a displayed polypeptide or protein on the surface of the minicell. The displayed polypeptide or protein may be a binding moiety directed to a compound or moiety displayed by a particular type of cell, or to a compound or moiety displayed by a pathogen. Such minicells can further, but need not, comprise an expression element having eukaryotic, archael, eucaryotic, or viral expression sequences operably linked to a nucleotide sequence that serves as a template for a bioactive nucleic acid.

[0032] In one aspect, the invention is drawn to immunogenic minicells, i.e., minicells that display an immunogen, vaccines comprising immunogenic minicells, antibodies and antibody derivatives directed to immunogens displayed on immunogenic minicells, and method of making and using immunogenic minicells and antibodies and antibody derivatives produced therefrom in prophylactic, diagnostic, therapeutic and research applications. A preferred immunogen displayed by a minicell is an immunogenic polypeptide, which is preferably expressed from an expression element contained within the minicell in order to maximize the amount of immunogen displayed by the immunogenic minicells. The immunogenic polypeptide can be derived from any organism, obligate intracellular parasite, organelle or virus with the proviso that, in prophylactic applications, the immunogenic polypeptide is not derived from a prokaryote, including a eubacterial virus. The source organism for the immunogen may be a pathogen. A minicell displaying an immunogen derived from a pathogen is formulated into a vaccine and, in a prophylactic application, used to treat or prevent diseases and disorders caused by or related to the eukaryotic or archebacterial pathogen.

[0033] In a separate aspect, the invention is drawn to minicells that display an immunogen derived from a nonfunctional, dysfunctional and/or diseased cell. By way of non-limiting example, the minicells display an immunogenic polypeptide derived from a hyperproliferative cell, i.e., a cell that is tumorigenic, or part of a tumor or cancer. As another non-limiting example, a cell that is infected with a virus or an obligate intracellular parasite (e.g., Rickettsiae) displays an immunogenic polypeptide that is encoded by the genome of the infected cell but is aberrantly expressed in an infected cell. A vaccine comprising a minicell displaying an immunogen derived from a nonfunctional, dysfunctional and/or diseased cell is used in methods of treating or preventing hyperproliferative diseases or disorders, including without limitation a cell comprising an intracellular pathogen.

[0034] In one aspect, the invention is drawn to methods of using minicells, and expression systems optimized therefore, to manufacture, on a large scale, proteins using recombinant DNA technology. In a related aspect, the invention is drawn to the production, via recombinant DNA technology, and/or segration of exogenous proteins in minicells. The minicells are enriched for the exogenous protein, which is desirable for increased yield and purity of the protein. In addition to protein purification, the minicells can be used for crystallography, the study of intracellular or extracellular protein-protein interactions, the study of intracellular or extracellular protein-nucleic acid interactions, the study of intracellular or extracellular protein-membrane interactions, and the study of other biological, chemical, or physiological event(s).

[0035] In one aspect, the invention is drawn to minicells having a membrane protein that has an intracellular domain.
By way of non-limiting example, the intracellular domain is exposed on the inner surface of the minicell membrane oriented towards the cytoplasmic compartment. The intracellular protein domain is available for interaction with intracellular components. Intracellular components may be naturally present in the minicells or their parent cells, or may be introduced into minicells after segregation from parent cells. A membrane-associated protein may have more than one intracellular domain, and a minicell of the invention may display more than one membrane-associated protein.

[0036] In one aspect, the invention is drawn to a minicell comprising a membrane protein that is linked to a conjugatable compound (a.k.a. “attachable compound”). The conjugatable compound may be of any chemical nature and have one or more therapeutic or detectable moieties. By way of non-limiting example, a protein having a transmembrane or membrane anchoring domain is displayed and has the capacity to be specifically cross-linked on its extracellular domain. Through this approach, any conjugatable compound of interest may be quickly and easily attached to the outer surface of minicells containing this expressed membrane-spanning domain. In aspects of the invention wherein minicells are used for drug delivery in vivo, a preferred conjugatable compound is polyethylene glycol (PEG), which provides for “stealth” minicells that are not taken as well and/or as quickly by the reticuloendothelial system (RES). Other conjugatable compounds include polysaccharides, polynucleotides, lipopolysaccharides, lipoproteins, glycosylated proteins, synthetic chemical compounds, and/or chimeric combinations of these examples listed.

[0037] In various aspects of the invention, the minicell displays a polypeptide or other compound or moiety on its surface. By way of non-limiting example, a non-eubacterial membrane protein displayed by eubacterial minicells may be a receptor. Minicells displaying a receptor may, but need not, bind ligands of the receptor. In therapeutic applications of this aspect of the invention, the ligand is an undesirable compound that is bound to its receptor and, in some aspects, is internalized by the minicells. In drug discovery applications of this aspect of the invention, the ligand for the receptor may be detectably labeled so that its binding to its receptor may be quantified. In the latter circumstance, the minicells may be used to identify and isolate, from a pool of compounds, one or more compounds that inhibit or stimulate the activity of the receptor. That is, these minicells can be used in screening assays, including assays such as those used in high-throughput screening (HTS) systems and other drug discovery methods, for the purpose of preparing compounds that influence the activity of a receptor of interest.

[0038] The non-eubacterial membrane protein displayed by minicells may be a fusion protein, i.e., a protein that comprises a first polypeptide having a first amino acid sequence and a second polypeptide having a second amino acid sequence, wherein the first and second amino acid sequences are not naturally present in the same polypeptide. At least one polypeptide in a membrane fusion protein is a “transmembrane domain” or “membrane-anchoring domain”. The transmembrane and membrane-anchoring domains of a membrane fusion protein may be selected from membrane proteins that naturally occur in a eukaryote, such as a fungus, a unicellular eukaryote, a plant and an animal, such as a mammal including a human. Such domains may be from a viral membrane protein naturally found in a virus such as a bacteriophage or a eukaryotic virus, e.g., an adenovirus or a retrovirus. Such domains may be from a membrane protein naturally found in an archaebacterium such as a thermophile. [0039] The displayed domain of a membrane fusion protein may be an enzymatic domain such as one having the activity of a lipase, a kinase, a phosphatase, a reductase, a protease, or a nuclease. Contacting such minicells with the appropriate substrate of the enzyme allows the substrate to be converted to reactant. When either the substrate or reactant is detectable, the reaction catalyzed by the membrane-bound enzyme may be quantified. In the latter instance, the minicells may be used to identify and isolate, from a pool of compounds, one or more compounds that inhibit or stimulate the activity of the enzyme represented by the displayed enzymatic moiety. That is, these minicells can be used in screening assays, including assays such as those used in high-throughput screening (HTS) systems and other drug discovery methods, for the purpose of preparing compounds that influence the activity of an enzyme or enzymatic moiety of interest.

[0040] The displayed domain of a membrane fusion protein may be a binding moiety. By way of non-limiting example, binding moieties used for particular purposes may be a binding moiety directed to a compound or moiety displayed by a specific cell type or cells found predominantly in one type of tissue, which may be used to target minicells and their contents to specific cell types or tissues; or a binding moiety that is directed to a compound or moiety displayed by a pathogen, which may be used in diagnostic or therapeutic methods; a binding moiety that is directed to an undesirable compound, such as a toxin, which may be used to bind and preferably internalize and/or neutralize the undesirable compound; a diseased cell; or the binding moiety may be a domain that allows for the minicells to be covalently or non-covalently attached to a support material, which may be used in compositions and methods for compound screening and drug discovery.

[0041] In one aspect, the invention provides compositions and methods for preparing a soluble and/or secreted protein where the protein remains in the cytoplasm of the minicell or is secreted following native secretory pathways for endogenous secreted proteins or is secreted using chimeric fusion to secretory signaling sequences. By way of non-limiting example, secreted or cytoplasmic soluble proteins may be produced for purification, targeted therapeutic applications where the protein produced is a therapeutic agent and is produced at the desired site of, detection for screening or diagnostic purposes where the protein is produced in response to a simulataneous and/or localization event, or to stimulate targeted minicell-cell fusion or interaction events where the protein produced stimulates cell-cell fusion upon targeted stimulation.

[0042] In one aspect, the invention provides compositions and methods for preparing antibodies and/or antibody derivatives that recognize an immunogenic epitope present on the native form of a membrane protein, but which is not immunogenic when the membrane protein is denatured or when prepared as a synthetic oligopeptide. Such antibodies and antibody derivatives are said to be “conformation sensitive.” Unlike most antibodies and antibody derivatives prepared by using a denatured membrane protein or an oligopeptide derived from the membrane protein, conformation sensitive antibodies and antibody derivatives specifically bind membrane proteins in their native state (i.e., in a membrane) with high affinity. Conformation sensitive antibodies and antibody derivatives are used to target compounds and compositions,
including a minicell of the invention, to a cell displaying the membrane protein of choice. Conformation sensitive antibodies and antibody derivatives are also used to prevent receptors from binding their natural ligands by specifically binding to the receptor with a high affinity and thereby limiting access of the ligand to the receptor. Conformation sensitive antibodies and antibody derivatives can be prepared that are specific for a specific isomer or mutant of a membrane protein, which can be useful in research and medical applications.

In one aspect, the invention provides biosensors comprising minicells including, not limited to, the minicells of the invention. An exemplary biosensor of the invention is a BIAcore chip, i.e., a chip onto which minicells are attached, where the minicells undergo some change upon exposure to a preselected compound, and the change is detected using surface plasmon resonance. A biosensor comprising minicells can be used in methods of detecting the presence of an undesirable compound. Undesirable compounds include but are not limited to, toxins; pollutants; explosives, such as those in landmines or illegally present: illegal narcotics; components of biological or chemical weapons. In a related aspect, the invention provides a device comprising a microchip operatively associated with a biosensor comprising a minicell. The device can further comprise an actuator that performs a responsive function when the sensor detects a preselected level of a marker.

In one aspect, the invention provides minicells that may be used as research tools and/or kits comprising such research tools. The minicells of the invention may be used as is, or incorporated into research tools useful for scientific research regarding all amino acid comprising compounds including, but not limited to, membrane-associated proteins, chimeric membrane fusion proteins, and soluble proteins. Such scientific research includes, by way of non-limiting example, basic research, as well as pharmacological, diagnostic, and pharmacogenetic studies. Such studies may be carried out in vivo or in vitro.

In one aspect, the invention is drawn to archaeobacterial minicells. In a related aspect, the invention is drawn to archaeobacterial minicells comprising at least one exogenous protein, that is, a protein that is not normally found in the parent cell, including without limitation fusion proteins. The archaeobacterial minicells of the invention optionally comprise an expression element that directs the production of the exogenous protein(s).

In other aspects, the invention is drawn to methods of preparing the minicells, protoplasts, and PoroplastsTM of the invention for various applications including but not limited to, diagnostic, therapeutic, research and screening applications. In a related aspect, the invention is drawn to pharmaceutical compositions, reagents and kits comprising minicells.

In each aspect and embodiment of the invention, unless stated otherwise, embodiments wherein the minicell is an archaeobacterial minicell, a poroplast, a spheroplast or a protoplast exist.

In a first aspect, the invention provides a minicell comprising a membrane protein selected from the group consisting of a eukaryotic membrane protein, an archaeobacterial membrane protein and an organellar membrane protein. In another embodiment, wherein the minicell comprises a biologically active compound. By way of non-limiting example, the biologically active compound is a radioisotope, a polypeptide, a nucleic acid or a small molecule.

In another embodiment, the minicell comprises a expression construct, wherein the first expression construct comprises expression sequences operably linked to an ORF that encodes a protein. In another embodiment, the ORF encodes the membrane protein. In another embodiment, the expression sequences that are operably linked to an ORF are inducible and/or repressible.

In another aspect, the minicell comprises a second expression construct, wherein the second expression construct comprises expression sequences operably linked to a gene. In another embodiment, the expression sequences that are operably linked to a gene are inducible and/or repressible. In a related embodiment, the gene product of the gene regulates the expression of the ORF that encodes the protein. A factor that "regulates" the expression of a gene or a gene product directly or indirectly initiates, enhances, quickens, slows, terminates, limits or completely blocks expression of a gene. In different embodiments, the gene product of the gene is a nucleic acid or a polypeptide. The polypeptide can be of any type, including but not limited to, a membrane protein, a soluble protein or a secreted protein. A membrane protein can be a membrane fusion protein comprising a first polypeptide, which comprises at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide.

In another aspect, the invention provides a minicell comprising a membrane fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising: at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide is not derived from a eubacterial protein and is neither a His tag nor a glutathione-S-transferase polypeptide. In various embodiments, the minicell is an eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the minicell comprises a biologically active compound.

In one aspect, the invention provides a minicell comprising a membrane conjugate, wherein the membrane conjugate comprises a membrane protein chemically linked to a conjugated compound. In one embodiment, the conjugated compound is selected from the group consisting of a nucleic acid, a polypeptide, a lipid and a small molecule.

In one aspect, the invention provides a method for making minicells, comprising: (a) culturing a minicell-producing parent cell, wherein the parent cell comprises an expression construct, wherein the expression construct comprises a gene operably linked to expression sequences that are inducible and/or repressible, and wherein induction or repression of the gene causes or enhances the production of minicells; and (b) separating the minicells from the parent cell, thereby generating a composition comprising minicells, wherein an inducer or repressor is present within the parent cells during one or more steps and/or between two or more steps of the method. In one embodiment, the method further comprises purifying the minicells from the composition.

Relevant gene products are factors involved in or modulating DNA replication, cellular division, cellular partitioning, septation, transcription, translation, or protein folding. The minicells are separated from parent cells by processes such as centrifugation, ultracentrifugation, density gradient, immunoaffinity, immunoprecipitation and other techniques described herein.
In one embodiment, the minicell is a poroplast, and the method further comprises (d) treating the minicells with an agent, or incubating the minicells under a set of conditions, that degrades the outer membrane of the minicell. The outer membrane is degraded by treatment with an agent selected from the group consisting of EDTA, EGTA, lactate acid, citric acid, gluconic acid, tartaric acid, polyethyleneimine, polycationic peptides, catiologic leucocyte peptides, aminoglucosides, aminoglycosides, proteamine, insect cecropins, reptilin maganins, polymers of basic amino acids, polymyxin B, chloroform, nitrolotriatic acid and sodium hexametaphosphate; by exposure to conditions selected from the group consisting of osmotic shock and sononation; and by other methods described herein.

In one embodiment, further comprising removing one or more contaminants from the composition. Representative contaminants are LPS and peptidoglycan. In a representative embodiment, LPS is removed by contacting the composition to an agent that binds or degrades LPS. At least about 50%, preferably about 65% to about 75%, more preferably 95%, most preferably 99% or >99% of LPS is removed from an initial preparation of minicells. In a related embodiment, the minicell-producing parent cell comprises a mutation in a gene required for lipopolysaccharide synthesis.

In one embodiment, the minicell is a spheroplast, and the method further comprises (d) treating the minicells with an agent, or incubating the minicells under a set of conditions, that disrupts or degrades the outer membrane; and (e) treating the minicells with an agent, or incubating the minicells under a set of conditions, that disrupts or degrades the cell wall. The agent that disrupts or degrades the cell wall can be, e.g., a lysozyme, and the set of conditions that disrupts or degrades the cell wall can be, e.g., incubation in a hypertonic solution.

In one embodiment, the minicell is a protoplast, and the method further comprises (d) treating the minicells with an agent, or incubating the minicells under a set of conditions, that disrupt or degrade the outer membrane; (e) treating the minicells with an agent, or incubating the minicells under a set of conditions, that disrupts or degrades the cell wall, in order to generate a composition that comprises protoplasts; and (f) purifying protoplasts from the composition. In one embodiment, the method further comprises preparing a denuded minicell from the minicell. In one embodiment, the method further comprises covalently or non-covalently linking one or more components of the minicell to a conjugated moiety.

In one aspect, the invention provides a L-form minicell comprising (a) culturing an L-form escherichia cell, wherein the escherichia cell comprises one or more of the following: (i) an expression element that comprises a gene operably linked to expression sequences that are inducible and/or repressible, wherein induction or repression of the gene regulates the copy number of an episomal expression construct; (ii) a mutation in an endogenous gene, wherein the mutation regulates the copy number of an episomal expression construct; (iii) an expression element that comprises a gene operably linked to expression sequences that are inducible and/or repressible, wherein induction or repression of the gene causes or enhances the production of minicells; and (iv) a mutation in an endogenous gene, wherein the mutation causes or enhances minicell production; (b) culturing the L-form minicell-producing parent cell in media under conditions wherein minicells are produced; and (c) separating the minicells from the parent cell, thereby generating a composition comprising L-form minicells, wherein an inducer or repressor is present within the minicells during one or more steps and/or between two or more steps of the method. In one embodiment, the method further comprises (d) purifying the L-form minicells from the composition.

In one aspect, the invention provides a method of producing a protein, comprising (a) transforming a minicell-producing parent cell with an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes the protein; (b) culturing the minicell-producing parent cell under conditions wherein minicells are produced; and (c) purifying minicells from the parent cell, (d) purifying the protein from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), and during step (c).

In one embodiment, the invention provides a method of producing a protein, comprising (a) transforming a minicell-producing parent cell with an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes the protein; (b) culturing the minicell-producing parent cell under conditions wherein minicells are produced; and (c) purifying minicells from the parent cell, (d) purifying the protein from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), and during step (c).

In one embodiment, the expression elements segregate into the minicells, and the ORF is expressed between steps (c) and (d). In one embodiment, the protein is a soluble protein contained within the minicells, and the method further comprises (e) lysing the minicells.

In one embodiment, the expression elements segregate into the minicells, and the method further comprises (e) collecting a composition in which the minicells are suspended or with which the minicells are in contact.

In one embodiment, the expression sequences to which the ORF is operably linked are inducible, wherein the method further comprises adding an inducing agent between steps (a) and (b); during step (b); and between steps (b) and (c).

In one embodiment, the expression sequences to which the ORF is operably linked are inducible, wherein the expression elements segregate into the minicells, the method further comprises adding an inducer to the minicell-containing parent cell. In one embodiment, the method further comprises (e) preparing protoplasts from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c); during step (c); between step (c) and step (d) when the expression elements segregate into the minicells; and/or after step (d) when the expression elements segregate into the minicells.

In one embodiment, the expression sequences to which the ORF is operably linked are inducible, wherein the method further comprises adding an inducing agent after step (c).

In one embodiment, the method further comprises (e) preparing protoplasts from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c); during step (c); between step (c) and step (d) when the expression elements segregate into the minicells; and/or after step (d) when the expression elements segregate into the minicells.

In one embodiment, the method further comprises (f) purifying the protein from the protoplasts.

In one embodiment, the method further comprises (e) purifying membrane preparations from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), during step (c), between steps (c) and (d) and/or after step (d).

In one embodiment, the method further comprises (f) purifying the protein from the protoplasts.

In one embodiment, the method further comprises (e) preparing protoplasts from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), during step (c), between steps (c) and (d) and/or after step (d).

In one embodiment, the method further comprises (f) purifying the protein from the membrane preparations.

In one embodiment, the method further comprises (e) preparing protoplasts from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), during step (c), between steps (c) and (d) and/or after step (d).

In one embodiment, the method further comprises (f) purifying the protein from the protoplasts.

In one embodiment, the method further comprises (e) preparing membrane preparations from the minicells, wherein the ORF is expressed during step (b), between steps (b) and (c), during step (c), between steps (c) and (d) and/or after step (d).

In one aspect, the invention provides a method of producing a protein, comprising (a) transforming a minicell
with an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes the protein; and (b) incubating the minicells under conditions wherein the ORF is expressed.

[0075] In one embodiment, the method further comprises (c) purifying the protein from the minicells.

[0076] In one aspect, the invention provides a method of producing a protein, comprising (a) transforming a minicell-producing parent cell with an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes a fusion protein comprising the protein and a polypeptide, wherein a protease-sensitive amino acid sequence is positioned between the protein and the polypeptide; (b) culturing the minicell-producing parent cell under conditions wherein minicells are produced; (c) purifying minicells from the parent cell, wherein the ORF is expressed during step (b); between steps (b) and (c); and/or after step (c) when the expression elements segregate into the minicells; and (d) treating the minicells with a protease that cleaves the sensitive amino acid sequence, thereby separating the protein from the polypeptide.

[0077] In one aspect, the invention provides a poroplast, the poroplast comprising a vesicle, bonded by a membrane, wherein the membrane is an eubacterial inner membrane, wherein the vesicle is surrounded by a eubacterial cell wall, and wherein the eubacterial inner membrane is accessible to a compound in solution with the poroplast. In one embodiment, the poroplast is a cellular poroplast. The compound has a molecular weight of at least 1 kDa, preferably at least about 0.1 to about 1 kDa, more preferably from about 1, 10 or 25 kDa to about 50 kDa, and most preferably from about 75 or about 100 kDa to about 150 or 300 kDa.

[0078] In one embodiment, the poroplast comprises an exogenous nucleic acid, which may be an expression construct. In one embodiment, the expression construct comprises an ORF that encodes an exogenous protein, wherein the ORF is operably linked to expression sequences. In one embodiment, the exogenous protein is a fusion protein, a soluble protein or a secreted protein. In one embodiment, the exogenous protein is a membrane protein, and is preferably accessible to compounds in solution with the poroplast. In one embodiment, poroplasts are placed in a hypertonic solution, wherein 90% or more of an equivalent amount of spheroplasts or protoplasts lyse in the solution under the same conditions.

[0079] In one embodiment, the membrane protein is selected from the group consisting of a eukaryotic membrane protein, an archeabacterial membrane protein, and an organellar membrane protein. In one embodiment, the membrane protein is a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide is displayed by the poroplast. In one embodiment, the second polypeptide is displayed on the external side of the eubacterial inner membrane. The second polypeptide can be an enzyme moiety, a binding moiety, a toxin, a cellular uptake sequence, an epitope, a detectable polypeptide, and a polypeptide comprising a conjugatable moiety. An enzyme moiety is a polypeptide derived from, by way of non-limiting example, a cytochrome P450, an oxidoreductase, a transferase, a hydrolase, a lyase, an isomerase, a ligase or a synthetase.

[0080] In one embodiment, the poroplast comprises a membrane component that is chemically linked to a conjugated compound.

[0081] In one embodiment, the expression construct comprises one or more DNA fragments from a genome or cDNA. In one embodiment, the exogenous protein has a primary amino acid sequence predicted from a nucleic acid sequence.

[0082] In one aspect, the invention provides a solid support comprising a minicell. In various embodiments, the solid support is a dipstick, a bead or a microtiter multwell plate. In one embodiment, the minicell comprises a detectable compound, which may be a colorimetric, fluorescent or radioactive compound.

[0083] In one embodiment, the minicell displays a membrane component selected from the group consisting of (i) a eukaryotic membrane protein, (ii) an archeabacterial membrane protein, (iii) an organellar membrane protein, (iv) a fusion protein comprising at least one transmembrane domain or at least one membrane anchoring domain, and (v) a membrane conjugate comprising a membrane component chemically linked to a conjugated compound.

[0084] In one embodiment, the membrane component is a receptor. In a related embodiment, the solid support further comprises a co-receptor. In one embodiment, the minicell displays a binding moiety.

[0085] In one aspect, the invention provides a solid support comprising a minicell, wherein the minicell displays a fusion protein, the fusion protein comprising a first polypeptide that comprises at least one transmembrane domain or at least one membrane anchoring domain, and a second polypeptide. In various embodiments, the second polypeptide comprises a binding moiety or an enzyme moiety.

[0086] In one aspect, the invention provides a solid support comprising a minicell, wherein the minicell comprises a membrane conjugate comprising a membrane component chemically linked to a conjugated compound. In one embodiment, the conjugated compound is a spacer. In one embodiment, the spacer is covalently linked to the solid support. In one embodiment, the conjugated compound is covalently linked to the solid support.

[0087] In one aspect, the invention provides a minicell comprising a biologically active compound, wherein the minicell displays a ligand or binding moiety, wherein the ligand or binding moiety is part of a fusion protein comprising a first polypeptide that comprises at least one transmembrane domain or at least one membrane anchoring domain and a second polypeptide that comprises a binding moiety, and the minicell is a poroplast, spheroplast or protoplast.

[0088] In one aspect, the invention provides a eubacterial minicell comprising a biologically active compound, wherein the minicell displays a binding moiety, wherein the binding moiety is selected from the group consisting of (a) a eukaryotic membrane protein; (b) an archeabacterial membrane protein; (c) an organellar membrane protein; and (d) a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide is not derived from a eubacterial protein and is neither a His tag nor a glutathione-S-transferase polypeptide, and wherein the polypeptide comprises a binding moiety.

[0089] In one embodiment, the binding moiety is selected from the group consisting of an antibody, an antibody derivative, a receptor and an active site of a non-catalytic derivative.
of an enzyme. In a preferred embodiment, the binding moiety is a single-chain antibody. In one embodiment, one of the ORFs encodes a protein that comprises the binding moiety.

[0090] In one embodiment, the binding moiety is directed to a ligand selected from the group consisting of an epitope displayed on a pathogen, an epitope displayed on an infected cell and an epitope displayed on a hyperproliferative cell.

[0091] In one embodiment, the invention further comprises a first and second nucleic acid, wherein the first nucleic acid comprises eukaryotic expression sequences operably linked to a first ORF, and a second nucleic acid, wherein the second nucleic acid comprises eubacterial expression sequences operably linked to a second ORF.

[0092] In one embodiment, the eubacterial expression sequences are induced and/or derepressed when the binding moiety is in contact with a target cell. In a variant embodiment, the eukaryotic expression sequences are induced and/or derepressed when the nucleic acid is in the cytoplasm of a eukaryotic cell. In related embodiments, the protein encoded by the first ORF comprises eukaryotic secretion sequences and/or the protein encoded by the second ORF comprises eubacterial secretion sequences.

[0093] In one aspect, the invention provides a method of associating a radioactive compound with a cell, wherein the cell displays a ligand specifically recognized by a binding moiety, comprising contacting the cell with a minicell that comprises the radioactive compound and displays the binding moiety. In a diagnostic embodiment, the amount of radiation emitted by the radioactive isotope is sufficient to be detectable. In a therapeutic embodiment, the amount of radiation emitted by the radioactive isotope is sufficient to be cytoxic.

In one embodiment, the ligand displayed by the cell is selected from the group consisting of an epitope displayed on a pathogen, an epitope displayed on an infected cell and an epitope displayed on a hyperproliferative cell. In one embodiment, the binding moiety is selected from the group consisting of an antibody, an antibody derivative, a channel protein and a receptor, and is preferably a single-chain antibody. In other embodiments, the binding moiety is an aptamer or a small molecule. In one embodiment, the ligand is selected from the group consisting of a cytokine, hormone, and a small molecule.

[0094] In one aspect, the invention provides a method of delivering a biologically active compound to a cell, wherein the cell displays a ligand specifically recognized by a binding moiety, comprising contacting the cell with a minicell that displays the binding moiety, wherein the minicell comprises the biologically active compound, and wherein the contents of the minicell are delivered into the cell from a minicell bound to the cell. In one embodiment, the biologically active compound is selected from the group consisting of a nucleic acid, a lipid, a polypeptide, a radioactive compound, an ion and a small molecule.

[0095] In one embodiment, the membrane of the minicell comprises a system for transferring a molecule from the interior of a minicell into the cytoplasm of the cell. A representative system for transferring a molecule from the interior of a minicell into the cytoplasm of the cell is a Type III secretion system.

[0096] In one embodiment, the minicell further comprises a first and second nucleic acid, wherein the first nucleic acid comprises eukaryotic expression sequences operably linked to a first ORF, and a second nucleic acid, wherein the second nucleic acid comprises eubacterial expression sequences operably linked to a second ORF. In one embodiment, one of the ORFs encodes a protein that comprises the binding moiety. In one embodiment, the eukaryotic expression sequences are induced and/or derepressed when the binding moiety is in contact with a target cell. In one embodiment, the eukaryotic expression sequences are induced and/or derepressed when the nucleic acid is in the cytoplasm of a eukaryotic cell. In one embodiment, the protein encoded by the first ORF comprises eukaryotic secretion sequences and/or the protein encoded by the second ORF comprises eubacterial secretion sequences.

In one embodiment, the ligand is selected from the group consisting of a cytokine, hormone, and a small molecule.

[0097] In one aspect, the invention provides a minicell displaying a synthetic linking moiety, wherein the synthetic linking moiety is covalently or non-covalently attached to a membrane component of the minicell.

[0098] In one aspect, the invention provides a sterically stabilized minicell comprising a displayed moiety that has a longer half-life in vivo than a wild-type minicell, wherein the displayed moiety is a hydrophilic polymer that comprises a PEG moiety, a carboxylic group of a polyalkylene glycol or PEG stearate.

[0099] In one aspect, the invention provides a minicell having a membrane comprising an exogenous lipid, wherein a minicell comprising the exogenous lipid has a longer half-life in vivo than a minicell lacking the exogenous lipid, and wherein the minicell is selected from the group consisting of a eubacterial minicell, a poroplast, a spheroplast and a proteoplast. In one embodiment, the exogenous lipid is a derivitized lipid which may, by way of non-limiting example, be phosphatidylethanolamine derivatized with PEG, DSPE-PEG, PEG stearate, PEG-derivatized phospholipids, a PEG ceramide or DSPE-PEG.

[0100] In one embodiment, the exogenous lipid is not present in a wild-type membrane, or is present in a different proportion than is found in minicells comprising a wild-type membrane. The exogenous lipid can be a ganglioside, sphingomyelin, monosialoganglioside GM1, galactocerebroside sulfate, 1,2-sn-dimyristoylphosphatidylethanolamine, phosphatidylglycerol, and cardiolipin.

[0101] In one embodiment, the linking moiety is non-covalently attached to the minicell. In one embodiment, one of the linking moiety and the membrane component comprises biotin, and the other comprises avidin or streptavidin. In one embodiment, the synthetic linking moiety is a cross-linker. In one embodiment, the cross-linker is a bifunctional cross-linker.

[0102] In one aspect, the invention provides a method of transferring a membrane protein from a minicell membrane to a biological membrane comprising contacting a minicell to the biological membrane, wherein the minicell membrane comprises the membrane protein, and allowing the minicell and the biological membrane to remain in contact for a period of time sufficient for the transfer to occur.

[0103] In one embodiment, the biological membrane is a cytoplasmic membrane or an organelar membrane. In one embodiment, the biological membrane is a membrane selected from the group consisting of a membranous two-layered cell and a membrane of a hyperproliferative cell. In one embodiment, the biological membrane is the cytoplasmic membrane of a recipient cell, which may be a cultured cell and a cell within an organism. In one embodiment, the biological membrane is present on a cell.
that has been removed from an animal, the contacting occurs in vitro, after which the cell is returned to the organism.

[0104] In one embodiment, the membrane protein is an enzyme. In this embodiment, the membrane protein having enzymatic activity is selected from the group consisting of a cytochrome P450 and a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one polypeptide, wherein the second polypeptide has enzymatic activity.

[0105] In one embodiment, the membrane protein is a membrane fusion protein, the membrane fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide.

[0106] In one embodiment, the second polypeptide is a biologically active polypeptide. In one embodiment, the minicell displays a ligand or a binding moiety.

[0107] In one aspect, the invention provides a minicell that comprises an expression construct comprising an ORF encoding a membrane protein operably linked to expression sequences, wherein the expression sequences are induced and/or derepressed when the minicell is in contact with a target cell.

[0108] In one embodiment, the biological membrane is a cytoplasmic membrane or an organellar membrane. In one embodiment, the biological membrane is a membrane selected from the group consisting of a membrane of a pathogen, a membrane of an infected cell and a membrane of a hyperproliferative cell. In one embodiment, the minicell displays a ligand or a binding moiety selected from the group consisting of an antibody, an antibody derivative, an aptamer and a small molecule. In one embodiment, the membrane protein is a membrane fusion protein, the membrane fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide. In one embodiment, the ligand is selected from the group consisting of a cytokine, hormone, and a small molecule.

[0109] In one aspect, the invention provides a pharmaceutical composition comprising a minicell, wherein the minicell displays a membrane protein, wherein the membrane protein is selected from the group consisting of a eukaryotic membrane protein, an archeabacterial membrane protein and an organellar membrane protein. In one embodiment, the membrane protein is selected from the group consisting of a receptor, a channel protein, a cellular adhesion factor and an integrin. In one embodiment, the pharmaceutical formulation further comprises an adjuvant. In one embodiment, the membrane protein comprises a polypeptide epitope displayed by a hyperproliferative cell. In one embodiment, the membrane protein comprises an epitope displayed by a eukaryotic pathogen, an archeabacterial pathogen, a virus or an infected cell.

[0111] In one aspect, the invention provides a pharmaceutical composition comprising a minicell, wherein the minicell displays a membrane conjugate, wherein the membrane conjugate comprises a membrane component chemically linked to a conjugated compound. In one embodiment, the membrane protein is selected from the group consisting of a receptor, a channel protein, a cellular adhesion factor and an integrin. In one embodiment, the pharmaceutical formulation further comprises an adjuvant. In one embodiment, the membrane protein is a polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain, or a lipid that is part of a membrane. In one embodiment, the conjugated compound is a polypeptide, and the chemical linkage between the membrane compound and the conjugated compound is not a peptide bond. In one embodiment, the conjugated compound is a nucleic acid. In one embodiment, the conjugated compound is an organic compound. In one embodiment, the organic compound is selected from the group consisting of a narcotic, a toxin, a venom, a sphingolipid and a soluble protein.

[0112] In one aspect, the invention provides a method of making a pharmaceutical composition comprising a minicell, wherein the minicell displays a membrane protein, wherein the membrane protein is selected from the group consisting of a eukaryotic membrane protein, an archeabacterial membrane protein and an organellar membrane protein. In one embodiment, the method further comprises adding an adjuvant to the pharmaceutical formulation. In one embodiment, the method further comprises desiccating the formulation. In one embodiment, the method further comprises adding a suspension buffer to the formulation. In one embodiment, the method further comprises making a chemical modification of the membrane protein. In one embodiment, the chemical modification is selected from the group consisting of glycosylation, deglycosylation, phosphorylation, dephosphorylation and proteinolysis. In one aspect, the invention provides a method of making a pharmaceutical composition comprising a minicell, wherein the minicell displays a membrane protein that is a fusion protein, the fusion protein comprising (i) a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and (ii) a second polypeptide, wherein the second polypeptide is not derived from a eubacterial protein.

[0113] In one aspect, the invention provides a method of making a pharmaceutical composition comprising a minicell, wherein the minicell displays a membrane conjugate, wherein the membrane conjugate comprises a membrane component chemically linked to a conjugated compound. In one embodiment, the method further comprises adding an adjuvant to the pharmaceutical formulation. In one embodiment, the membrane component is a polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain, or a lipid that is part of a membrane. In one embodiment, the conjugated compound is a polypeptide, and the chemical linkage between the membrane compound and the conjugated compound is not a peptide bond. In one embodiment, the conjugated compound is a nucleic acid. In one embodiment, the conjugated compound is an organic compound. In one embodiment, the organic compound is selected from the group consisting of a narcotic, a toxin, a venom, and a sphingolipid.
In one aspect, the invention provides a method of detecting an agent that is specifically bound by a binding moiety, comprising contacting a minicell displaying the binding moiety with a composition known or suspected to contain the agent, and detecting a signal that is modulated by the binding of the agent to the binding moiety. In one embodiment, the agent is associated with a disease. In one embodiment, the minicell comprises a detectable compound. In one embodiment, the binding moiety is antibody or antibody derivative. In one embodiment, the composition is an environmental sample. In one embodiment, the composition is a biological sample. In one embodiment, the biological sample is selected from the group consisting of blood, serum, plasma, urine, saliva, a biopsy sample, feces and a skin patch.

In one aspect, the invention provides a method of in situ imaging of a tissue or organ, comprising administering to an organism a minicell comprising an imaging agent and a binding moiety and detecting the imaging agent in the organism.

In one embodiment, the minicell is a eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the binding moiety is an antibody or antibody derivative. In one embodiment, the binding moiety specifically binds a cell surface antigen. In one embodiment, the cell surface antigen is an antigen displayed by a tumorigenic cell, a cancer cell, and an infected cell. In one embodiment, the cell surface antigen is a tissue-specific antigen. In one embodiment, the method of imaging is selected from the group consisting of magnetic resonance imaging, ultrasound imaging and computer axial tomography (CAT). In one aspect, the invention provides a device comprising a microchip operatively associated with a biosensor comprising a minicell, wherein the microchip comprises or contacts the minicell, and wherein the minicell displays a binding moiety.

In one embodiment, the invention provides a method of detecting a substance that is specifically bound by a binding moiety, comprising contacting the device of claim 16 with a composition known or suspected to contain the substance, and detecting a signal from the device, wherein the signal changes as a function of the amount of the substance present in the composition. In one embodiment, the composition is a biological sample or an environmental sample.

In one aspect, the invention provides a method of identifying an agent that specifically binds a target compound, comprising contacting a minicell displaying the target compound with a library of compounds, and identifying an agent in the library that binds the target compound. In one embodiment, the library of compounds is a protein library. In one embodiment, the protein library is selected from the group consisting of a phage display library, a phagemid display library, a baculovirus library, a yeast display library, and a ribosomal display library. In one embodiment, the library of compounds is selected from the group consisting of a library of aptamers, a library of synthetic peptides and a library of small molecules.

In one embodiment, the target compound is a target polypeptide. In one embodiment, the minicell comprises an expression construct comprising expression sequences operably linked to an ORF encoding the target polypeptide. In one embodiment, the target polypeptide is a membrane protein. In one embodiment, the membrane protein is a receptor or a channel protein. In one embodiment, the membrane protein is an enzyme. In one embodiment, the target compound is a membrane fusion protein, the membrane fusion protein comprising a first polypeptide, wherein the first polypeptide comprises at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide comprises amino acid sequences derived from a target polypeptide. In one embodiment, the method further comprises comparing the activity of the target compound in the presence of the agent to the activity of the target compound in the absence of the agent.

In one embodiment, the activity of the target compound is an enzyme activity. In one embodiment, the activity of the target compound is a binding activity. In one embodiment, the invention further comprises comparing the binding of the agent to the target compound to the binding of a known ligand of the target compound. In one embodiment, a competition assay is used for the comparing.

In one aspect, the invention provides a device comprising microchips operatively associated with a biosensor comprising a set of minicells in a prearranged pattern, wherein the each of the microchips comprise or contact a minicell, wherein each of the minicells displays a different target compound, and wherein binding of a ligand to a target compound results in an increased or decreased signal. In one embodiment, the invention provides a method of identifying an agent that specifically binds a target compound, comprising contacting the device with a library of compounds, and detecting a signal from the device, wherein the signal changes as a function of the binding of an agent to the target compound. In one embodiment, the invention provides a method of identifying an agent that specifically blocks the binding of a target compound to its ligand, comprising contacting the device with a library of compounds, and detecting a signal from the device, wherein the signal changes as a function of the binding of an agent to the target compound.

In one aspect, the invention provides a method of making an antibody that specifically binds a protein domain, wherein the domain is in its native conformation, wherein the domain is contained within a protein displayed on a minicell, comprising contacting the minicell with a cell, wherein the cell is competent for producing antibodies to an antigen contacted with the cell, in order to generate an immunogenic response in which the cell produces the antibody.

In one embodiment, the protein displayed on a minicell is a membrane protein. In one embodiment, the membrane protein is a receptor or a channel protein. In one embodiment, the domain is found within the second polypeptide of a membrane fusion protein, wherein the membrane fusion protein comprises a first polypeptide, wherein the first polypeptide comprises at least one transmembrane domain or at least one membrane anchoring domain. In one embodiment, the contacting occurs in vivo. In one embodiment, the antibody is a polyclonal antibody or a monoclonal antibody. In one embodiment, the contacting occurs in an animal that comprises an adjuvant.

In one aspect, the invention provides the method of making an antibody derivative that specifically binds a protein domain, wherein the domain is in its native conformation, wherein the domain is displayed on a minicell, comprising contacting the minicell with a protein library, and identifying an antibody derivative from the protein library that specifically binds the protein domain. In one embodiment, the protein library is selected from the group consisting of a phage display library, a phagemid display library, and a ribosomal display library.
In one aspect, the invention provides a method of making an antibody or antibody derivative that specifically binds an epitope, wherein the epitope is selected from the group consisting of (i) an epitope composed of amino acids found within a membrane protein, (ii) an epitope present in an interface between a membrane protein and a membrane component, (iii) an epitope present in an interface between a membrane protein and one or more other proteins and (iv) an epitope in a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain, and a second polypeptide, the second polypeptide comprising the epitope; comprising contacting a minicell displaying the epitope with a protein library, or to a cell, wherein the cell is competent for producing antibodies to an antigen contacted with the cell, in order to generate an immunogenic response in which the cell produces the antibody.

In one embodiment, the cell is contacted in vivo. In various embodiments, the antibody is a polyclonal antibody or a monoclonal antibody. In one embodiment, the protein library is contacted in vitro. In one embodiment, the protein library is selected from the group consisting of a phage display library, a plasmid display library, and a ribosomal display library.

The invention provides a method of determining the rate of transfer of a nucleic acid from a minicell to a cell, comprising (a) contacting the cell to the minicell, wherein the minicell comprises the nucleic acid, for a measured period of time; (b) separating minicells from the cells; (c) measuring the amount of nucleic acid in the cells, wherein the amount of nucleic acid in the cells over the set period of time is the rate of transfer of a nucleic acid from a minicell.

In one aspect, the invention provides a method of determining the amount of a nucleic acid transferred to a cell from a minicell, comprising (a) contacting the cell to the minicell, wherein the minicell comprises an expression element having eukaryotic expression sequences operably linked to an ORF encoding a detectable polypeptide, wherein the minicell displays a binding moiety, and wherein the binding moiety binds an epitope of the cell; and (b) detecting a signal from the detectable polypeptide, wherein a change in the signal corresponds to an increase in the amount of a nucleic acid transferred to a cell.

In one embodiment, the cell is a eukaryotic cell. By way of non-limiting example, a eukaryotic cell can be a plant cell, a fungal cell, a unicellular eukaryote, an animal cell, a mammalian cell, a rat cell, a mouse cell, a primate cell or a human cell.

In one embodiment, the binding moiety is an antibody or antibody derivative. In one embodiment, the binding moiety is a single-chain antibody. In one embodiment, the binding moiety is an aptamer. In one embodiment, the binding moiety is an organic compound.

In one aspect, the invention provides a method of detecting the expression of an expression element in a cell, comprising (a) contacting the cell to a minicell, wherein the minicell comprises an expression element having expression sequences operably linked to an ORF encoding a detectable polypeptide, wherein the minicell displays a binding moiety, and wherein the binding moiety binds an epitope of the cell; (b) incubating the cell and the minicell for a period of time effective for transfer of nucleic acid from the minicell to the cell; and (c) detecting a signal from the detectable polypeptide, wherein an increase in the signal corresponds to an increase in the expression of the expression element.

In one embodiment, the cell is a eukaryotic cell and the expression sequences are eukaryotic expression sequences. In one embodiment, the eukaryotic cell is a mammalian cell. In one embodiment, the binding moiety is an antibody or antibody derivative. In one embodiment, the binding moiety is a single-chain antibody. In one embodiment, the binding moiety is an aptamer. In one embodiment, the binding moiety is an organic compound.

In a related aspect, the invention provides methods of detecting the transfer of a fusion protein from the cytosol to an organelle of a eukaryotic cell, comprising (a) contacting the cell to a minicell, wherein (i) the minicell comprises an expression element having eukaryotic expression sequences operably linked to an ORF encoding a fusion protein, wherein the fusion protein comprises a first polypeptide that comprises organelar delivery sequences, and a second polypeptide that comprises a detectable polypeptide; and (ii) the minicell displays a binding moiety that binds an epitope of the cell, or an epitope of an organelle; (b) incubating the cell and the minicell for a period of time effective for transfer of nucleic acid from the minicell to the cell and production of the fusion protein; and (c) detecting a signal from the detectable polypeptide, wherein a change in the signal corresponds to an increase in the amount of the fusion protein transferred to the organelle.

In one aspect, the invention provides a minicell comprising at least one nucleic acid, wherein the minicell displays a binding moiety directed to a target compound, wherein the binding moiety is selected from the group consisting of (i) a eukaryotic membrane protein; (ii) an archaeobacterial membrane protein; (iii) an organelar membrane protein; and (iv) a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide is not derived from a eubacterial protein and is not a His tag nor a glutathione-S-transferase polypeptide, and wherein the polypeptide comprises a binding moiety.

In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF encoding a protein selected from the group consisting of (i) the eukaryotic membrane protein, (ii) the archaeobacterial membrane protein, (iii) the organelar membrane protein; and (iv) the fusion protein.

In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF, wherein the ORF encodes a therapeutic polypeptide. In one embodiment, the therapeutic polypeptide is a membrane polypeptide. In one embodiment, the therapeutic polypeptide is a soluble polypeptide. In one embodiment, the soluble polypeptide comprises a cellular secretion sequence. In one embodiment, the expression sequences are inducible and/or repressible.

In one embodiment, the expression sequences are induced and/or derepressed when the binding moiety displayed by the minicell binds to its target compound. In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF, wherein the ORF encodes a polypeptide having an amino acid sequence that facilitates cellular transfer of a biologically active compound contained within or displayed
by the minicell. In one embodiment, the membrane of the minicell comprises a system for transferring a molecule from the interior of a minicell into the cytoplasm of the cell. In one embodiment, the system for transferring a molecule from the interior of a minicell into the cytoplasm of the cell is a Type III secretion system.

[0138] In one aspect, the invention provides a method of introducing a nucleic acid into a cell, comprising contacting the cell with a minicell that comprises the nucleic acid, wherein the minicell displays a binding moiety, wherein the binding moiety is selected from the group consisting of (i) a eukaryotic membrane protein; (ii) an archaebacterial membrane protein; (iii) an organellar membrane protein; and (iv) a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain; and a second polypeptide, wherein the second polypeptide is not derived from a eubacterial protein and is neither a His tag nor a glutathione-S-transferase polypeptide, and wherein the polypeptide comprises a binding moiety, and wherein the binding moiety binds an epitope of the cell.

[0139] In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF encoding a protein selected from the group consisting of (i) the eukaryotic membrane protein, (ii) the archaebacterial membrane protein, (iii) the organellar membrane protein; and (iv) a fusion protein.

[0140] In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF, wherein the ORF encodes a therapeutic polypeptide. In one embodiment, the expression sequences are inducible and/or derepressible. In one embodiment, the expression sequences are induced or derepressed when the binding moiety displayed by the minicell binds its target compound. In one embodiment, the expression sequences are induced or derepressed by a transactivation or transrepression event. In one embodiment, the nucleic acid comprises an expression construct comprising expression sequences operably linked to an ORF, wherein the ORF encodes a polypeptide having an amino acid sequence that facilitates cellular transfer of a biologically active compound contained within or displayed by the minicell.

[0141] In one aspect, the invention provides a minicell comprising a nucleic acid, wherein the nucleic acid comprises eukaryotic expression sequences and eubacterial expression sequences, each of which is independently operably linked to an ORF.

[0142] In one embodiment, the minicell displays a binding moiety. In one embodiment, the eubacterial expression sequences are induced and/or derepressed when the binding moiety is in contact with a target cell. In one embodiment, the eukaryotic expression sequences are induced and/or derepressed when the nucleic acid is in the cytoplasm of a eukaryotic cell. In one embodiment, the protein encoded by the ORF comprises eubacterial or eukaryotic secretion sequences.

[0143] In one aspect, the invention provides a minicell comprising a first and second nucleic acid, wherein the first nucleic acid comprises eukaryotic expression sequences operably linked to a first ORF, and a second nucleic acid, wherein the second nucleic acid comprises eubacterial expression sequences operably linked to a second ORF.

[0144] In one embodiment, the minicell displays a binding moiety. In one embodiment, the eubacterial expression sequences are induced and/or derepressed when the binding moiety is in contact with a target cell. In one embodiment, the eukaryotic expression sequences are induced and/or derepressed when the nucleic acid is in the cytoplasm of a eukaryotic cell. In one embodiment, the protein encoded by the first ORF comprises eukaryotic secretion sequences and/or the protein encoded by the second ORF comprises eubacterial secretion sequences.

[0145] In one aspect, the invention provides a method of introducing into and expressing a nucleic acid in an organism, comprising contacting a minicell to a cell of the organism, wherein the minicell comprises a nucleic acid.

[0146] In one embodiment, the minicell displays a binding moiety. In one embodiment, the nucleic acid comprises a eukaryotic expression construct, wherein the eukaryotic expression construct comprises eukaryotic expression sequences operably linked to an ORF. In one embodiment, the ORF encodes a protein selected from the group consisting of a membrane protein, a soluble protein and a protein comprising eukaryotic secretion signal sequences. In one embodiment, the nucleic acid comprises a eubacterial expression construct, wherein the eubacterial expression construct comprises eubacterial expression sequences operably linked to an ORF. In one embodiment, the ORF encodes a protein selected from the group consisting of a membrane protein, a soluble protein and a protein comprising eukaryotic secretion signal sequences. In one aspect, the invention provides a minicell comprising a crystal of a membrane protein. In one embodiment, the minicell is a eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the membrane protein is a receptor. In one embodiment, the receptor is a G-protein coupled receptor. In one embodiment, the crystal is displayed.

[0147] In a related aspect, the invention provides a minicell membrane preparation comprising a crystal of a membrane protein.

[0148] In one embodiment, the membrane protein is a fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain, and a second polypeptide. In one embodiment, the crystal is a crystal of the second polypeptide. In one embodiment, the crystal is displayed.

[0149] In one aspect, the invention provides a method of determining the three-dimensional structure of a membrane protein, comprising preparing a crystal of the membrane protein in a minicell, and determining the three-dimensional structure of the crystal.

[0150] In one aspect, the invention provides a method for identifying ligand-interacting atoms in a defined three-dimensional structure of a target protein, comprising (a) preparing one or more variant proteins of a target protein having a known or predicted three-dimensional structure, wherein the target protein binds a preselected ligand; (b) expressing and displaying a variant protein in a minicell; and (c) determining if a minicell displaying the variant protein binds the preselected ligand with increased or decreased affinity as compared to the binding of the preselected ligand to the target protein.

[0151] In one embodiment, the ligand is a protein that forms a multimer with the target protein, and the ligand interacting atoms are atoms in the defined three-dimensional structure that are involved in protein-protein inter-
actions. In one embodiment, the ligand is a compound that induces a conformational change in the target protein, and the defined three-dimensional structure is the site of the conformational change. In one embodiment, the method for identifying ligands of a target protein, further comprising identifying the chemical differences in the variant proteins as compared to the target protein. In one embodiment, the invention further comprises mapping the chemical differences onto the defined three-dimensional structure, and correlating the effect of the chemical differences on the defined three-dimensional structure. In one embodiment, the target protein is a wild-type protein. In one aspect, the invention provides a minicell library, comprising two or more minicells, wherein each minicell comprises a different exogenous protein. In one embodiment, the minicell is a eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the exogenous protein is a displayed protein. In one embodiment, the exogenous protein is a membrane protein. In one embodiment, the membrane protein is a receptor. In one embodiment, the protein is a soluble protein that is contained within or secreted from the minicell. In one embodiment, minicells within the library comprise an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes the exogenous protein. In one embodiment, the nucleic acid has been mutagenized; the mutagenesis can be site-directed or random. In one embodiment, an active site of the exogenous protein has a known or predicted three-dimensional structure, and the portion of the ORF encoding the active site has been mutagenized. In one embodiment, each of the minicells comprises an exogenous protein that is a variant of a protein having a known or predicted three-dimensional structure.

[0152] In one aspect, the invention provides a minicell library, comprising two or more minicells, wherein each minicell comprises a different fusion protein, each of the fusion protein comprising a first polypeptide that is a constant polypeptide, wherein the constant polypeptide comprises at least one transmembrane domain or at least one membrane anchoring domain, and a second polypeptide, wherein the second polypeptide is a variable amino acid sequence that is different in each fusion proteins. In one embodiment, minicells within the library comprise an expression element that comprises expression sequences operably linked to a nucleic acid having an ORF that encodes the fusion protein. In one embodiment, the second polypeptide of the fusion protein is encoded by a nucleic acid that has been cloned. In one embodiment, each of the second polypeptide of each of the fusion proteins comprises a variant of an amino acid sequence from a protein having a known or predicted three-dimensional structure.

[0153] In one aspect, the invention provides a minicell library, comprising two or more minicells, wherein each minicell comprises a constant protein that is present in each minicell and a variable protein that differs from minicell to minicell. In one embodiment, one of the constant and variable proteins is a receptor, and the other of the constant and variable proteins is a co-receptor. In one embodiment, each of the constant and variable proteins is different from each other and is a factor in a signal transduction pathway. In one embodiment, one of the constant and variable proteins is a G-protein, and the other of the constant and variable proteins is a G-protein coupled receptor.

[0154] In one embodiment, one of the constant and variable proteins comprises a first transrepression domain, and the other of the constant and variable proteins comprises a second transrepression domain, wherein the transrepression domains limit or block expression of a reporter gene when the constant and variable proteins associate with each other.

[0155] In one embodiment, one of the constant and variable proteins comprises a first transactivation domain, and the other of the constant and variable proteins comprises a second transactivation domain, wherein the transactivation domains stimulate expression of a reporter gene when the constant and variable proteins associate with each other.

[0156] In one aspect, the invention provides a method of identifying a nucleic acid that encodes a protein that binds to or chemically alters a preselected ligand, comprising: (a) separately contacting the ligand with individual members of a minicell library, wherein minicells in the library comprise expression elements, wherein the expression elements comprise DNA inserts, wherein an ORF in the DNA insert is operably linked to expression sequences, in order to generate a series of reaction mixes, each reaction mix comprising a different member of the minicell library; (b) incubating the reaction mixes, thereby allowing a protein that binds to or chemically alters the preselected ligand to bind or chemically alter the preselected ligand; (c) detecting a change in a signal from reaction mixes in which the ligand has been bound or chemically altered; (d) preparing DNA from reaction mixes in which the ligand has been bound or chemically altered; wherein the DNA is a nucleic acid that encodes a protein that binds to or chemically alters the preselected ligand.

[0157] In one embodiment, the minicell is a eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the preselected ligand is a biologically active compound. In one embodiment, the preselected ligand is a therapeutic drug. In one embodiment, a protein that binds or chemically alters the preselected ligand is a target protein for compounds that are therapeutic for a disease that is treated by administering the drug to an organism in need thereof. In one embodiment, the preselected ligand is detectably labeled, the minicell comprises a detectable compound, and/or a chemically altered derivative of the protein is detectably labeled.

[0158] In one aspect, the invention provides a method of determining the amino acid sequence of a protein that binds or chemically alters a preselected ligand, comprising: (a) contacting the ligand with a minicell library, wherein minicells in the library comprise expression elements, wherein the expression elements comprise DNA inserts, wherein an ORF in the DNA insert is operably linked to expression sequences; (b) incubating the mixture of ligand and minicells, under conditions which allow complexes comprising ligands and minicells to form and/or chemical reactions to occur; (c) isolating or identifying the complexes from the ligand and the mixture of ligand and minicells; (d) preparing DNA from an expression element found in one or more of the complexes, or in a minicell thereof; (e) determining the nucleotide sequence of the ORF in the DNA; and (f) generating an amino sequence by in silico translation, wherein the amino acid sequence is or is derived from a protein that binds or chemically alters a preselected ligand.

[0159] In one embodiment, the minicell is a eubacterial minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the DNA is prepared by isolating DNA from the complexes, or in a minicell thereof. In one embodiment, the DNA is prepared by amplifying DNA from the complexes, or in a minicell thereof. In one embodiment, the protein is a fusion protein. In one embodiment, the protein is a membrane...
or a soluble protein. In one embodiment, the protein comprises secretion sequences. In one embodiment, the preselected ligand is a biologically active compound. In one embodiment, the preselected ligand is a therapeutic drug. In one embodiment, the preselected ligand is a therapeutic drug, and the protein that binds the preselected ligand is a target protein for compounds that are therapeutic for a disease that is treated by administering the drug to an organism in need thereof.

[0160] In one aspect, the invention provides a method of identifying a nucleic acid that encodes a protein that inhibits or blocks an agent from binding to or chemically altering a preselected ligand, comprising: (a) separately contacting the ligand with individual members of a minicell library, wherein minicells in the library comprise expression elements, wherein the expression elements comprise DNA inserts, wherein an ORF in the DNA insert is operably linked to expression sequences, in order to generate a series of reaction mixes, each reaction mix comprising a different member of the minicell library; (b) incubating the reaction mixes, thereby allowing a protein that binds to or chemically alters the preselected ligand to bind to or chemically alter the preselected ligand; (c) detecting a change in a signal from reaction mixes in which the ligand has been bound or chemically altered; (d) preparing DNA from reaction mixes in which the change in signal ligand has been bound or chemically altered; wherein the DNA is a nucleic acid that encodes a protein that inhibits or blocks the agent from binding to or chemically altering the preselected ligand.

[0161] In one embodiment, the minicell is a subcellular minicell, a poroplast, a spheroplast or a protoplast. In one embodiment, the DNA has a nucleotide sequence that encodes the amino acid sequence of the protein that inhibits or blocks the agent from binding to or chemically altering the preselected ligand. In one embodiment, a protein that binds or chemically alters the preselected ligand is a target protein for compounds that are therapeutic for a disease that is treated by administering the drug to an organism in need thereof.

[0162] In one aspect, the invention provides a method of identifying an agent that affects the activity of a protein, comprising contacting a library of two or more candidate agents with a minicell comprising the protein or a polypeptide derived from the protein, assaying the effect of candidate agents on the activity of the protein, and identifying agents that affect the activity of the protein.

[0163] In one embodiment, the protein or the polypeptide derived from the protein is displayed on the surface of the minicell. In one embodiment, the protein is a membrane protein. In one embodiment, the membrane protein is selected from the group consisting of a receptor, a channel protein and an enzyme. In one embodiment, the activity of a protein is a binding activity or an enzymatic activity. In one embodiment, the library of compounds is a protein library. In one embodiment, the protein library is selected from the group consisting of a phage display library, a phagemid display library, and a ribosomal display library. In one embodiment, the library of compounds is a library of aptamers. In one embodiment, the library of compounds is a library of small molecules.

[0164] In one aspect, the invention provides a method of identifying an agent that affects the activity of a protein domain containing a library of two or more candidate agents with a minicell displaying a membrane fusion protein, the fusion protein comprising a first polypeptide, the first polypeptide comprising at least one transmembrane domain or at least one membrane anchoring domain, and a second polypeptide, wherein the second polypeptide comprises the protein domain.

[0165] In one aspect, the invention provides a method of identifying undesirable side-effects of a biologically active compound that occur as a result of binding of the compound to a protein, wherein binding a compound to the protein is known to result in undesirable side effects, comprising contacting a minicell that comprises the protein to the biologically active compound. In one embodiment, the invention provides comprises characterizing the binding of the biologically active compound to the protein. In one embodiment, the invention provides comprises characterizing the effect of the biologically active compound on the activity of the protein.

[0166] In one aspect, the invention provides a method for identifying an agent that effects the interaction of a first signaling protein with a second signaling protein, comprising (a) contacting a library of compounds with a minicell, wherein the minicell comprises: (i) a first protein comprising the first signaling protein and a first trans-acting regulatory domain; (ii) a second protein comprising the second signaling protein and a second trans-acting regulatory domain; and (iii) a reporter gene, the expression of which is modulated by the interaction between the first trans-acting regulatory domain and the second trans-acting regulatory domain; and (b) detecting the gene product of the reporter gene.

[0167] In one embodiment, the trans-acting regulatory domains are transactivation domains. In one embodiment, the trans-acting regulatory domains are transrepression domains.

[0168] In one embodiment, the reporter gene is induced by the interaction of the first trans-acting regulatory domain and the second trans-acting regulatory domain. In one embodiment, the agent that effects the interaction of the first signaling protein with the second signaling protein is an agent that causes or promotes the interaction. In one embodiment, the reporter gene is repressed by the interaction of the first trans-acting regulatory domain and the second trans-acting regulatory domain. In one embodiment, the agent that effects the interaction of the first signaling protein with the second signaling protein is an agent that inhibits or blocks the interaction.

[0169] In one embodiment, the first signaling protein is a GPCR. In one embodiment, the GPCR is an Edg receptor or a ScAMPER.

[0170] In one embodiment, the second signalling protein is a G-protein. In related embodiments, G-protein is selected from the group consisting of G-alpha-i, G-alpha-s, G-alpha-q, G-alpha-12/13 and Go. In one embodiment, the library of compounds is a protein library. In one embodiment, the protein library is selected from the group consisting of a phage display library, a phagemid display library, and a ribosomal display library. In one embodiment, the library of compounds is a library of aptamers. In one embodiment, the library of compounds is a library of small molecules.

[0171] In one aspect, the invention provides a method for identifying an agent that effects the interaction of a first signaling protein with a second signaling protein, comprising contacting a library of two or more candidate agents with a minicell, wherein the minicell comprises (a) a first fusion protein comprising the first signaling protein and a first detectable domain; and (b) a second fusion protein comprising the second signaling protein and a second detectable
domain, wherein a signal is generated when the first and second signaling proteins are in close proximity to each other, and detecting the signal.

[0172] In one embodiment, the signal is fluorescence. In one embodiment, the first detectable domain and the second detectable domain are fluorescent and the signal is generated by FRET. In one embodiment, the first and second detectable domains are independently selected from the group consisting of a green fluorescent protein, a blue-shifted green fluorescent protein, a cyan-shifted green fluorescent protein; a red-shifted green fluorescent protein; a yellow-shifted green fluorescent protein; and a red fluorescent protein, wherein the first fluorescent domain and the second fluorescent domain are not identical.

[0173] In one aspect, the invention provides a method of bioremediation, the method comprising contacting a composition that comprises an undesirable substance with a mini-cell, wherein the mini-cell alters the chemical structure and/or binds the undesirable substance.

[0174] In one aspect, the invention provides a method of bioremediation, the method comprising contacting a composition that comprises an undesirable substance with a mini-cell, wherein the mini-cell comprises an agent that alters the chemical structure of the undesirable substance. In one embodiment, the agent that alters the chemical structure of the undesirable substance is an inorganic catalyst. In one embodiment, the agent that alters the chemical structure of the undesirable substance is an enzyme. In one embodiment, the enzyme is a soluble protein contained within the mini-cell. In one embodiment, the enzyme is a secreted protein. In one embodiment, the enzyme is a membrane protein. In one embodiment, the membrane enzyme is selected from the group consisting of a cytochrome P450, an oxidoreductase, a transferase, a hydrolase, a lyase, an isomerase, a ligase and a synthetase. In one embodiment, the agent that alters the chemical structure of the undesirable substance is a fusion protein comprising a first polypeptide that comprises a transmembrane domain or at least one membrane-anchoring domain, and a second polypeptide, wherein the second polypeptide is an enzyme moiety.

[0175] In one aspect, the invention provides a method of bioremediation, the method comprising contacting a composition that comprises an undesirable substance with a mini-cell, wherein the mini-cell comprises an agent that binds an undesirable substance. In one embodiment, the undesirable substance binds to and is internalized by the mini-cell or is otherwise inactivated by selective absorption. In one embodiment, the agent that binds the undesirable substance is a secreted soluble protein. In one embodiment, the secreted protein is a transport accessory protein. In one embodiment, the agent that binds the undesirable substance is a membrane protein. In one embodiment, the undesirable substance is selected from the group consisting of a toxin, a pollutant and a pathogen. In one embodiment, the agent that binds the undesirable substance is a fusion protein comprising a first polypeptide that comprises a transmembrane domain or at least one membrane-anchoring domain, and a second polypeptide, wherein the second polypeptide is a binding moiety.

[0176] In one aspect, the invention provides a mini-cell producing parent cell, wherein the parent cell comprises one or more of the following (a) an expression element that comprises a gene operably linked to expression sequences that are inducible and/or repressible, wherein induction or repression of the gene regulates the copy number of an episomal expression construct; (b) a mutation in an endogenous gene, wherein the mutation regulates the copy number of an episomal expression construct; (c) an expression element that comprises a gene operably linked to expression sequences that are inducible and/or repressible, wherein induction or repression of the gene causes or enhances the production of mini-cells; and (d) a mutation in an endogenous gene, wherein the mutation causes or enhances mini-cell production.

[0177] In one embodiment, the invention comprises an episomal expression construct. In one embodiment, the invention further comprises a chromosomal expression construct. In one embodiment, the expression sequences of the expression construct are inducible and/or repressible. In one embodiment, the mini-cell-producing parent cell comprises a biologically active compound. In one embodiment, the gene that causes or enhances the production of mini-cells has a gene product that is involved in or regulates DNA replication, cellular division, cellular partitioning, septation, transcription, translation, or protein folding.

[0178] In one aspect, the invention provides a mini-cell producing parent cell, wherein the parent cell comprises an expression construct, wherein the expression construct comprises expression sequences operably linked to an ORF that encodes a protein, and a regulatory expression element, wherein the regulatory expression element comprises expression sequences operably linked to a regulatory gene that encodes a factor that regulates the expression of the ORF. In one embodiment, the expression sequences of the expression construct are inducible and/or repressible. In one embodiment, the expression sequences of the regulatory expression construct are inducible and/or repressible. In one embodiment, one or more of the expression element or the regulatory expression element is located on a chromosome of the parent cell. In one embodiment, one or more of the expression element or the regulatory expression element is located on an episomal expression construct. In one embodiment, both of the expression element and the regulatory expression element are located on an episomal expression construct, and one or both of the expression element and the regulatory expression element segregates into mini-cells produced from the parent cell. In one embodiment, the mini-cell producing parent cell comprises a biologically active compound. In one embodiment, the biologically active compound segregates into mini-cells produced from the parent cell. In one embodiment, the ORF encodes a membrane protein or a soluble protein. In one embodiment, the protein comprises secretion sequences. In one embodiment, the gene product of the gene regulates the expression of the ORF. In one embodiment, the gene product is a transcription factor. In one embodiment, the gene product is a RNA polymerase. In one embodiment, the parent cell is MC-T7.

[0179] In one aspect, the invention provides a mini-cell comprising a biologically active compound, wherein the mini-cell displays a binding moiety, wherein the mini-cell selectively absorbs and/or internalizes an undesirable compound, and the mini-cell is a poroplast, spheroplast or protoplast. In one embodiment, the binding moiety is selected from the group consisting of an antibody, an antibody derivative, the active site of a non-enzymatically active mutant enzyme, a single-chain antibody and an aptamer.
chain antibody. In one embodiment, the binding moiety is directed to a ligand selected from the group consisting of an epitope displayed on a pathogen, an epitope displayed on an infected cell and an epitope displayed on a hyperproliferative cell. In one embodiment, the biologically active compound is selected from the group consisting of a radioisotope, a polypeptide, a nucleic acid and a small molecule. In one embodiment, a ligand binds to and is internalized by the minicell or is otherwise inactivated by selective absorption. In one embodiment, the invention provides a pharmaceutical composition comprising the minicell. In one aspect, the invention provides a method of reducing the free concentration of a substance in a composition, wherein the substance displays a ligand specifically recognized by a binding moiety, comprising contacting the composition with a minicell that displays the binding moiety, wherein the binding moiety binds the substance, thereby reducing the free concentration of the substance in the composition. In one embodiment, the substance is selected from the group consisting of a nucleic acid, a lipid, a polypeptide, a radiolabeled compound, an ion and a small molecule. In one embodiment, the binding moiety is selected from the group consisting of an antibody, an antibody derivative, a channel protein and a receptor.

[0180] In one embodiment, the composition is present in an environment including but not limited to water, air or soil. In one embodiment, the composition is a biological sample from an organism, including but not limited to blood, serum, plasma, urine, saliva, a biopsy sample, feces, tissue and a skin patch. In one embodiment, the substance binds to and is internalized by the minicell or is otherwise inactivated by selective absorption. In one embodiment, the biological sample is returned to the organism after being contacting to the minicell.

[0181] For a better understanding of the present invention, reference is made to the accompanying detailed description and its scope will be pointed out in the appended claims. All references cited herein are hereby incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0182] FIG. 1 is a Western blot in which Edg-1-6xHis and Edg-3-6xHis proteins expressed in minicells produced from MC-17 cells.

[0183] FIG. 2 shows induction of Malle(L)-NTR in isolated minicells.

ABBREVIATIONS AND DEFINITIONS

[0184] For brevity’s sake, the single-letter amino acid abbreviations are used in some instances herein. Table 1 describes the correspondence between the 1- and 3-letter amino acid abbreviations.

<table>
<thead>
<tr>
<th>TABLE 1-continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>THREE- AND ONE-LETTER ABBREVIATIONS FOR AMINO ACIDS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Three-letter abbreviation</th>
<th>One-letter symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Glu</td>
<td>Q</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gyl</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>

[0185] A “conjugatable compound” or “attachable compound” is capable of being attached to another compound. The terms “conjugated to” and “cross-linked with” indicate that the conjugatable compound is in the state of being attached to another compound. A “conjugate” is the compound formed by the attachment of a conjugatable compound or conjugatable moiety to another compound.

[0186] “Culturing” signifies incubating a cell or organism under conditions wherein the cell or organism can carry out some, if not all, biological processes. For example, a cell that is cultured may be growing or reproducing, or it may be non-viable but still capable of carrying out biological and/or biochemical processes such as replication, transcription, translation, etc.

[0187] An agent is said to have been “purified” if its concentration is increased, and/or the concentration of one or more undesirable contaminants is decreased, in a composition relative to the composition from which the agent has been purified. Purification thus encompasses enrichment of an agent in a composition and/or isolation of an agent therefrom.

[0188] A “solid support” is any solid or semisolid composition to which an agent can be attached or contained within. Common forms of solid support include, but are not limited to, plates, tubes, and beads, all of which could be made of glass or another suitable material, e.g., polystyrene, nylon, cellulose acetate, nitrocellulose, and other polymers. Semisolids and gels that minicells are suspended within are also considered to be solid supports. A solid support can be in the form of a dipstick, flow-through device, or other suitable configuration.

[0189] A “mutation” is a change in the nucleotide sequence of a gene relative to the sequence of the “wild-type” gene. Reference wild-type entobacterial strains are those that have been cultured in vitro by scientists for decades; for example, a wild-type strain of Escherichia coli is E. coli K-12. Mutations include, but are not limited to, point mutations, deletions, insertions and translocations.

[0190] A “trans-acting regulatory domain” is a regulatory part of a protein that is expressed from a gene that is not adjacent to the site of regulatory effect. Trans-acting domains can activate or stimulate (transactivate), or limit or block (transrepress) the gene in question.

[0191] A “reporter gene” refers to a gene that is operably linked to expression sequences, and which expresses a gene
product, typically a detectable polypeptide, the production and detection of which is used as a measure of the robustness and/or control of expression.

[0192] A “detectable compound” or “detectable moiety” produces a signal that can be detected by spectroscopic, photochemical, biochemical, immunochemical, electromagnetic, radiophysical, or chemical means such as fluorescence, chemiluminescence, or chemiluminescence response, or any other appropriate means. A “radioactive compound” or “radioactive composition” has more than the natural (environmental) amount of one or more radioisotopes.

[0193] By “displayed” it is meant that a portion of the membrane protein is present on the surface of a cell or minicell, and is thus in contact with the external environment of the cell or minicell. The external, displayed portion of a membrane protein is an “extracellular domain” or a “displayed domain.” A membrane protein may have more than one displayed domain, and a minicell of the invention may display more than one membrane protein.

[0194] A “domain” or “protein domain” is a region of a molecule or structure that shares common physical and/or chemical features. Non-limiting examples of protein domains include hydrophobic transmembrane or peripheral membrane binding regions, globular enzymatic or receptor regions, and/or nucleic acid binding domains.

[0195] A “transmembrane domain” spans a membrane, a “membrane anchoring domain” is positioned within, but does not traverse, a membrane. An “extracellular” or “displayed” domain is present on the exterior of a cell, or minicell, and is thus in contact with the external environment of the cell or minicell.

[0196] A “eukaryote” is as the term is used in the art. A eukaryote may, by way of example, be a fungus, a unicellular eukaryote, a plant or an animal. An animal may be a mammal, such as a rat, a mouse, a rabbit, a dog, a cat, a horse, a cow, a pig, a simian or a human.

[0197] A “eukaryotic membrane” is a membrane found in a eukaryote. A eukaryotic membrane may, by way of example, a cytoplasmic membrane, a nuclear membrane, a cytosol, nucleolus, a membrane of the endoplasmic reticulum (ER), a membrane of a Golgi body, a membrane of a lysosome or a peroxisome, a caveolar membrane, or an inner or outer membrane of a mitochondrial, chloroplast, plastid or vacuole.

[0198] The terms “endogenous” refers to something that is normally found in a cell as that cell exists in nature.

[0199] The term “exogenous” refers to something that is not normally found in a cell as that cell exists in nature.

[0200] A “gene” comprises (a) nucleotide sequences that either (i) act as a template for a nucleic acid gene product, or (ii) that encode one or more open reading frames (ORFs); and (b) expression sequences operably linked to (1) or (2). When a gene comprises an ORF, it is a “structural gene.”

[0201] By “immunogenic,” it is meant that a compound elicits production of antibodies or antibody derivatives and, additionally or alternatively, a T-cell mediated response, directed to the compound or a portion thereof. The compound is an “immunogen.”

[0202] A “ligand” is a compound, composition or moiety that is capable of specifically bound by a binding moiety, including without limitation, a receptor and an antibody or antibody derivative.

[0203] A “membrane protein” is a protein found in whole or in part in a membrane. Typically, a membrane protein has (1) at least one membrane anchoring domain, (2) at least one transmembrane domain, or (3) at least one domain that interacts with a protein having (1) or (2).

[0204] An “ORF” or “open reading frame” is a nucleotide sequence that encodes an amino acid sequence of a known, predicted or hypothetical polypeptide. An ORF is bounded on its 5′ end by a start codon (usually ATG) and on its 3′ end by a stop codon (i.e., TAA or TAG). An ORF encoding a 20 amino acid sequence comprises 33 nucleotides (for each of 10 amino acids and 3 for a stop codon). ORFs can encode amino acid sequences that comprise from 10, 25, 50, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or more amino acids.

[0205] The terms “Eubacteria” and “prokaryote” are used herein as these terms are used by those in the art. The terms “eubacterial” and “prokaryotic” encompasses Eubacteria, including both gram-negative and gram-positive bacteria, prokaryotic viruses (e.g., bacteriophage), and obligate intracellular parasites (e.g., Rickettsia, Chlamydia, etc.).

[0206] An “active site” is any portion or region of a molecule required for, or that regulates, an activity of the molecule. In the case of a protein, an active site can be a binding site for a ligand or a substrate, an active site of enzyme, a site that directs or undergoes conformational change in response to a signal, or a site of post-translational modification of a protein.

[0207] In a poroplast, the eubacterial outer membrane (OM) and LPS have been removed. In a poroplast, portions of a disrupted eubacterial OM and/or disrupted cell wall either may remain associated with the inner membrane of the minicell, but the membrane is nonetheless porous because the permeability of the disrupted OM has been increased. A membrane is the to be “disrupted” when the membrane’s structure has been treated with an agent, or incubated under conditions, that leads to the partial degradation of the membrane, thereby increasing the permeability thereof. In contrast, a membrane that has been “degraded” is essentially, for the applicable intents and purposes, removed. In preferred embodiments, irrespective of the condition of the OM and cell wall, the eubacterial inner membrane is not disrupted, and membrane proteins displayed on the inner membrane are accessible to compounds that are brought into contact with the minicell, poroplast, spheroplast, proteoplast or cellular poroplast, as the case may be.

[0208] Host cells (and/or minicells) harboring an expression construct are components of expression systems.

[0209] An “expression vector” is an artificial nucleic acid molecule into which an exogenous ORF encoding a protein, or a template of a bioactive nucleic acid can be inserted in such a manner so as to be operably linked to appropriate expression sequences that direct the expression of the exogenous gene. Preferred expression vectors are episomal vectors that can replicate independently of chromosomal replication.

[0210] By the term “operably linked” it is meant that the gene products encoded by the non-vector nucleic acid sequences are produced from an expression element in vivo.

[0211] The term “gene product” refers to either a nucleic acid (the product of transcription, reverse transcription, or replication) or a polypeptide (the product of translation) that is produced using the non-vector nucleic acid sequences as a template.

[0212] An “expression construct” is an expression vector into which a nucleotide sequence of interest has been inserted.
in a manner so as to be positioned to be operably linked to the expression sequences present in the expression vector. Preferred expression constructs are episomal.

[0213] An “expression element” is a nucleic acid having nucleotide sequences that are present in an expression construct but not its cognate expression vector. That is, an expression element for a polypeptide is a nucleic acid that comprises an ORF operably linked to appropriate expression sequences. An expression element can be removed from its expression construct and placed in other expression vectors or into chromosomal DNA.

[0214] “Expression sequences” are nucleic acid sequences that bind factors necessary for the expression of genes that have been inserted into an expression vector. An example of an expression sequence is a promoter, a sequence that binds RNA polymerase, which is the enzyme that produces RNA molecules using DNA as a template. An example of an expression sequence that is both inducible and repressible is the L-arabinose operon (araC). See Schleif R. Regulation of the L-arabinose operon of Escherichia coli. Trends Genet. 2000 December; 16(12):559-65.

[0215] In the present disclosure, “a nucleic acid” or “the nucleic acid” refers to a specific nucleic acid molecule. In contrast, the term “nucleic acid” refers to any collection of diverse nucleic acid molecules, and thus signifies that any number of different types of nucleic acids are present. By way of non-limiting example, a nucleic acid may be a DNA, a dsRNA, a tRNA (including a rare codon usage tRNA), an mRNA, a ribosomal RNA (rRNA), a peptide nucleic acid (PNA), a DNA:RNA hybrid, an antisense oligonucleotide, a ribozyme, or an aptamer.

DETAILED DESCRIPTION OF THE INVENTION

[0216] The invention described herein is drawn to compositions and methods for the production of achromosomal archeabacterial, euabacterial and anucleate eukaryotic cells that are used for diagnostic and therapeutic applications, for drug discovery, and as research tools.

[0217] The general advantage of minicells over cell-based expression systems (e.g., eucaryotic cells or bacterial expression systems) is that one may express heterologous membrane bound proteins or over express endogenous membrane bound proteins, cytoplasmic or secreted soluble proteins, or small molecules on the cytoplasmic or extracellular surfaces of the minicells that would otherwise be toxic to live cells. Minicells are also advantageous for proteins that require a particular lipid environment for proper functioning because it is very manipulatable in nature. Other advantages include the stability of the minicells due to the lack of toxicity, the high level of expression that can be achieved in the minicell, and the efficient flexible nature of the minicell expression system. Such minicells could be used for in vivo targeting or for selective absorption (i.e., molecular “sponges”) and that these molecules can be expressed and “displayed” at high levels. Minicells can also be used to display proteins for low, medium, high, and ultra high throughput screening, crystal formation for structure determination, and for in vitro research use only applications such as transfection. Minicells expressing proteins or small molecules, radioisotopes, image-enhancing reagents can be used for in vivo diagnostics and for in vitro diagnostic and assay platforms. Also, soluble and/or membrane associated signaling cascade elements may be reconstituted in minicells producing encapsulated devices to follow extracellular stimulation events using cytoplasmic reporter events, e.g., transactivation resulting from dimerization of dimerization dependent transcriptional activation or repression of said reporter.

[0218] Regarding protein expression, minicells can be engineered to express one or more recombinant proteins in order to produce more protein per surface area of the particle (at least 10x more protein per unit surface area of protein). The proteins or small molecules that are “displayed” on the minicell surfaces have therapeutic, discovery or diagnostic benefit either when injected into a patient or used in a selective absorption mode during dillization. In vitro assays include drug screening and discovery, structural proteomics, and other functional proteomics applications. Proteins that are normally soluble can be tethered to membrane anchoring domains or membrane proteins can be expressed for the purpose of displaying these proteins on the surfaces of the minicell particle in therapeutic, discovery, and diagnostic modes. The types of proteins that can be displayed include but are not limited to receptors (e.g., GPCRs, sphingolipid receptors, neurotransmitter receptors, sensory receptors, growth factor receptors, hormone receptors, chemokine receptors, cytokine receptors, immunological receptors, and complement receptors, FC receptors), channels (e.g., potassium channels, sodium channels, calcium channels), pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potassium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidase), enzymes and kinases (e.g., protein kinases, ATPases, GTPases, phosphatases, proteases), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN), chemotactic/adhesion proteins (e.g., ICAM1, selectins, CD34, VCAM-1, LFA-1, VLA-1), and chimeric/fusion proteins (e.g., proteins in which a normally soluble protein is attached to a transmembrane region of another protein). As a non-limiting example, the small molecules that can be tethered and displayed on the surfaces of the minicells can be carbohydrates (e.g., monosaccharides), bioactive lipids (e.g., lysosphingolipids, PAF, lysophospholipids), drugs (e.g., antibiotics, ion channel activators/inhibitors, ligands for receptors and/or enzymes), nucleic acids (e.g., synthetic oligonucleotides), fluorophores, metals, or inorganic and organic small molecules typically found in combinatorial chemistry libraries. Minicells may either contain (encapsulate) or display on their surfaces radioisotides or image-enhancing reagents both of which could be used for therapeutic and/or diagnostic benefit in vivo or for in vitro assays and diagnostic platforms.

[0219] For in vivo therapeutic uses, minicells can express proteins and/or display small molecules on their surfaces that would either promote an immune response and passage through the RES system (e.g., to eliminate the minicell and its target quickly), or to evade the RES (e.g., to increase the bioavailability of the minicell). Toxicity is reduced or eliminated because the therapeutic agent is not excreted or processed by the liver and thus does not damage the kidneys or liver, because the minicell-based therapeutic is not activated until entry into the target cell (e.g., in the case of cancer therapeutics or gene therapy). Minicells are of the appropriate size (from about 0.005, 0.1, 0.15 or 0.2 micrometers to about 0.25, 0.3, 0.35, 0.4, 0.45 or 0.5 micrometers) to facilitate deep penetration into the lungs in the cases where administration of the minicell-based therapeutic or diagnostic is via an inhalant (Strong, A. A., et al. 1987. An aerosol generator system for
inhalation delivery of pharmacological agents. Med. Instrum. 21:189-194). This is due to the fact that minicells can be aerosolized. Without being limited to the following examples, inhalant therapeutic uses of minicells could be applied to the treatment of anaphylactic shock, viral infection, inflammatory reactions, gene therapy for cystic fibrosis, treatment of lung cancers, and fetal distress syndrome.

Minicells can also display expressed proteins that are enzymes that may have therapeutic and/or diagnostic uses. The enzymes that are displayed may be soluble enzymes that are expressed as fusion proteins with a transmembrane domain of another protein. Display of such enzymes could be used for in vitro assays or for therapeutic benefit.

Gene therapy applications afforded by minicells generally involve the ability of minicells to deliver DNA to target cells (either for replacement therapy, modification of cell function or to kill cells). Expression plasmids can be delivered to target cells that would encode proteins that could be cytoplasmic or could have intracellular signal sequences that would target the protein to a particular organelle (e.g., mitochondria, nuclei, endoplasmic reticulum, etc.). In the case where minicells are engulfed by the target cell, the minicells themselves could have those intracellular targeting sequences expressed on their surfaces so that the minicells could be "delivered" to intracellular targets.

Minicells used for the following therapeutic, discovery, and diagnostic applications can be prepared as described in this application and then stored and/or packaged by a variety of ways, including but not limited to lyophilization, freezing, mixing with preservatives (e.g., antioxidants, glycerol), or otherwise stored and packaged in a fashion similar to methods used for liposome and protoliposome formulations.

The small size of minicells (from about 0.005, 0.1, 0.15 or 0.2 micrometers to about 0.25, 0.3, 0.35, 0.4, 0.45 or 0.5 micrometers) makes them suitable for many in vitro diagnostic platforms, including the non-limiting examples of lateral flow, ELISA, HTS, especially those applications requiring microspheres or nanospheres that display many target proteins or other molecules. The use of proteolipids or poroplast minicells may be especially useful in this regard. Assay techniques are dependent on cell or particle size, protein (or molecule to be tested) amount displayed on the surface of the cell or particle, and the sensitivity of the assay being measured. In current whole-cell systems, the expression of the protein of interest is limiting, resulting in the higher cell number requirement to satisfy the sensitivity of most assays. However, the relatively large size of cells prevents the incorporation of large numbers of cells in these assays, e.g., 96, 384, and smaller well formats. In contrast, minicells, proteolipids, and poroplasts are smaller in size and can be manipulated to express high levels of the presellected protein, and can be incorporated into small well assay formats.

Types of Minicells

Minicells are derivatives of cells that lack chromosomal DNA and which are sometimes referred to as amucleate cells. Because eubacterial and archaeabacterial cells, unlike eukaryotic cells, do not have a nucleus (a distinct organelle that contains chromosomes), these non-eukaryotic minicells are more accurately described as being "without chromosomes" or "achromosomal," as opposed to "amucleate." Nonetheless, those skilled in the art often use the term "amucleate" when referring to bacterial minicells in addition to other minicells. Accordingly, in the present disclosure, the term "minicells" encompasses derivatives of eubacterial cells that lack a chromosome; derivatives of archaeabacterial cells that lack their chromosom(e)s (Laurence et al., Nucleoid Structure and Partition in Methanococcus jannaschii: An Archaeon With Multiple Copies of the Chromosome, Genetics 152:1315-1323, 1999); and amucleate derivatives of eukaryotic cells. It is understood, however, that some of the relevant art may use the terms "amucleate minicells" or "amucleate cells" loosely to refer to any of the preceding types of minicells.

Eubacterial Minicells


Eukaryotic Minicells

The term “eukaryote” is defined as is used in the art, and includes any organism classified as Eucarya that are usually classified into four kingdoms: plants, animals, fungi and protists. The first three of these correspond to phylogenetically coherent groups. However, the eukaryotic protists do not form a group, but rather are comprised of many phylogenetically disparate groups (including slime molds, multiple groups of algae, and many distinct groups of protozoa). See, e.g., Olsen, G., http://www.bact.wisc.edu/microtextbook/. A type of animal of particular interest is a mammal, including, by way of non-limiting example a rat, a mouse, a rabbit, a dog, a cat, a horse, a cow, a pig, a simian and a human.

Chromosomeless eukaryotic minicells (i.e., amucleate cells) are within the scope of the invention. Platelets are a non-limiting example of eukaryotic minicells. Platelets are amucleate cells with little or no capacity for de novo protein synthesis. The tight regulation of protein synthesis in platelets (Smith et al., Platelets and stroke, Vasc Med 4:165-72, 1999) may allow for the over-production of exogenous proteins and, at the same time, under-production of endogenous proteins. Thrombin-activated expression elements such as those that
are associated with Bel-3 (Weyrich et al., Signal-dependent translation of a regulatory protein, Bel-3, in activated human platelets, Cell Biology 95:5556-5561, 1998) may be used to modulate the expression of exogenous genes in platelets.

As another non-limiting example, eukaryotic minicells are generated from tumor cell lines (Gygongossy-Issa and Khachaturians, Tumour minicells: single, large vesicles released from cultured mastocytoma cells (1985) Tissue Cell 17:801-809; Melton, Cell Fusion-induced mouse neuroblastomas HPRT revertants with variant enzyme and elevated HPRT protein levels (1981) Somatic Cell Genet 7: 331-344).

Yeast cells are used to generate fungal minicells. See, e.g., Lee et al., Ibp1p, a possible spindle pole body associated protein, regulates nuclear division and bud separation in Saccharomyces cerevisiae, Biochim Biophys Acta 3:239-253, 1999; Kopecka et al., A method of isolating anucleated yeast protoplasts unable to synthesize the glucan fibrillar component of the wall J Gen Microbiol 81:111-120, 1974; and Yoo et al., Fission yeast Hrp1, a chromodomain ATPase, is required for proper chromosome segregation and its overexpression interferes with chromatin condensation, Nucl Acids Res 28:2004-2011, 2000. Cell division in yeast is reviewed by Gould and Simanis, The control of septum formation in fission yeast, Genes & Dev 11:2939-51, 1997).

I.C. Archeabacterial Minicells

The term “archaeabacterium” is defined as in used in the art and includes extreme thermophiles and other Archaea. Woese, C. R., L. Magrum, G. Fox. 1978. Archeabacteria. Journal of Molecular Evolution. 11:245-252. Three types of Archeabacteria are halophiles, thermophiles and methanogens. By physiological definition, the Archaea (informally, archaebacteria) are single-cell extreme thermophiles (including thermococci, Halobacterium halobium, and Methanococcus halophilus), sulfate reducers, methanogens, and extreme halophiles. The thermophilic members of the Archaea include the most thermophilic organisms cultivated in the laboratory. The aerobic thermophiles are also acidophilic; they oxidize sulfur in their environment to sulfuric acid. The extreme halophiles are aerobic or microaerophilic and include the most salt tolerant organisms known. The sulfate-reducing Archaea reduce sulfate to sulfide in extreme environment. Methanogens are strict anaerobes, yet they gave rise to at least two separate aerobic groups: the halophiles and a thermococcal line (Olsen, G., http://www.bact. wisc.edu/microtextbook/). Non-limiting examples of halophiles include Halobacterium cutirubrum and Halorhages mediterranei. Non-limiting examples of methanogens include Methanococcus voltae; Methanococcus vannielii; Methanobacterium thermoautotrophicum; Methanococcus voltae; Methanothermus fervidus; and Methanosarcina barkeri. Non-limiting examples of thermophiles include Azobacter vinelandii; Thermoplasma acidophilum; Pyrococcus furiosus; and Crenarchaeota (extremely thermophile archaebacteria) species such as Sulfolobus solfataricus and Sulfolobus acidocaldarius.

Archeabacterial minicells are within the scope of the invention. Archeabacteria have homologs of eubacterial minicell genes and proteins, such as the MinD polypeptide from Pyrococcus furiosus (Hayashi et al., EMBO J. 2001 20:1819-28, Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus). It is thus possible to create Archeabacterial minicells by methods such as, by way of non-limiting example, overexpressing the product of a min gene isolated from a prokaryote or an archeabacterium; or by disrupting expression of a min gene in an archaebacterium of interest by, e.g., the introduction of mutations thereof or antisense molecules thetore. See, e.g., Laurence et al., Nucleoid Structure and Partition in Methanococcus jannaschii: An Archaeon With Multiple Copies of the Chromosome, Genetics 152:1315-1323, 1999.

In one aspect, the invention is drawn to archaebacterial minicells. By physiological definition, the Archaea (informally, archaebacteria) are single-cell extreme thermophiles (including thermococci), sulfate reducers, methanogens, and extreme halophiles. The thermophilic members of the Archaea include the most thermophilic organisms cultivated in the laboratory. The aerobic thermophiles are also acidophilic; they oxidize sulfur in their environment to sulfuric acid. The extreme halophiles are aerobic or microaerophilic and include the most salt tolerant organisms known. The sulfate-reducing Archaea reduce sulfate to sulfide in extreme environment. Methanogens are strict anaerobes, yet they gave rise to at least two separate aerobic groups: the halophiles and a thermococcal line (Olsen, G., http://www.bact. wisc.edu/microtextbook/).

I.D. Minicells Produced from Diverse Organisms

There are genes that can be disrupted to cause minicell production that are conserved among the three Kingdoms. For example, SMC (structural maintenance of chromosomes) proteins are conserved among prokaryotes, archaebacteria and eukaryotes (Hirano, SMC-mediated chromosome and mechanics: a conserved scheme from bacteria to vertebrates?; Genes and Dev. 13:11-19, 1999; Holmes et al., Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling, PNAS 97:1322-1324, 2000; Michiko and Hirano, EMBO J 17:7139-7148, 1998, ATP-dependent aggregation of single-stranded DNA by a bacterial SMC homodimer, 1998). Mutations in B. subtilis smc genes result in the production of minicells (Britton et al., Characterization of a eubacterial SMC protein involved in chromosome partitioning, Genes and Dev. 12:1254-1259, 1998; Moriya et al., A Bacillus subtilis gene encoding protein homologous to eukaryotic SMC motor protein is necessary for chromosome partition Mol Microbiol 29:179-87, 1998). Disruption of smc genes in various cells is predicted to result in minicell production therefrom.

As another example, mutations in the yeast genes encoding TRF topoisomerases result in the production of minicells, and a human homolog of yeast TRF genes has been stated to exist (Castano et al., A novel family of TRF (DNA topoisomerase I-related function) genes required for proper nuclear segregation, Nucleic Acids Res 24:2404-10, 1996). Mutations in a yeast chromodomain ATPase, Hrp1, result in abnormal chromosomal segregation; (Yoo et al., “Fission yeast Hrp1, a chromomain ATPase, is required for proper chromosome segregation and its overexpression interferes with chromatin condensation,” Nuc. Acids Res. 28:2004-2001). Disruption of TRF and/or Hrp1 function is predicted to cause minicell production in various cells. Genes involved in septum formation in fission yeast (see, e.g., Gould et al., “The control of septum formation in fission yeast,” Genes and Dev. 11:2939-2951, 1997) can be used in like fashion.

As another example, mutations in the divIVA gene of Bacillus subtilis results in minicell production (Table 2). When expressed in E. coli or the yeast Schizosaccharomyces pombe, a B. subtilis DivIVA-GFP protein is targeted to cell division sites therein, even though clear homologs of DivIVA do not seem to exist in E. coli or S. pombe (Dovid et al.,
Promiscuous targeting of Bacillus subtilis cell division protein DivIVA to division sites in Escherichia coli and fission yeast, EMBO J. 19:2719-2727, 2000.) Over- or under-expression of B. subtilis DivIVA or a homolog thereof may be used to reduce minicell production in a variety of cells.

[0241] II. Production of Minicells

[0242] Eubacterial minicells are produced by parent cells having a mutation in, and/or overexpressing, or under expressing a gene involved in cell division and/or chromosomal partitioning, or from parent cells that have been exposed to certain conditions, that result in aberrant fission of bacterial cells and/or partitioning in abnormal chromosomal segregation during cellular fission (division). The term “parent cells” or “parental cells” refers to the cells from which minicells are produced. Minicells, most of which lack chromosomal DNA (Mulder et al., The Escherichia coli miniB mutation resembles gyrB in Defective nucleoid segregation and decreased negative supercoiling of plasmids. Mol Gen Genet, 1990, 221: 87-93), are generally, but need not be, smaller than their parent cells. Typically, minicells produced from E. coli cells are generally spherical in shape and are about 0.1 to about 0.5 um in diameter, whereas whole E. coli cells are about from about 1 to about 3 um in diameter and from about 2 to about 10 um in length. Micrographs of E. coli cells and minicells that have been stained with DAPI (4:6-diamidino-2-phenylindole), a compound that binds to DNA, show that the minicells do not stain while the parent E. coli are brightly stained. Such micrographs demonstrate the lack of chromosomal DNA in minicells. (Mulder et al., Mol. Gen. Genet. 221:87-93, 1990).

[0243] As shown in Table 2, minicells are produced by several different mechanisms as such, by way of non-limiting example, the over expression of genes involved in chromosomal replication and partitioning, mutations in such genes, and exposure to various environmental conditions. “Overexpression” refers to the expression of a polypeptide or protein encoded by a DNA introduced into a host cell, wherein the polypeptide or protein is either not normally present in the host cell, or wherein the polypeptide or protein is present in the host cell at a higher level than that normally expressed from the endogenous gene encoding the polypeptide or protein. For example, in E. coli cells that overexpress the gene product FtsZ (The FtsZ gene encodes a protein that is involved in regulation of divisions; see Cook and Rothfield, Early stages in development of the Escherichia coli cell division site. Mol Microbiol., 1994, 14: p. 485-495; and Lutkenhaus, Regulation of cell division in E. coli. Trends Genet, 1990, 6: p. 232-253), there is an increase in the formation of minicells (Begg et al., Roles of FtsA and FtsZ in the activation of division sites. J. Bacteriology, 1997. 180: 881-884). Minicells are also produced by E. coli cells having a mutation in one or more genes of the min locus, which is a group of genes that encode proteins that are involved in cell division (de Boer et al., Central role for the Escherichia coli miniC gene product in two different cell division-inhibition systems. Proc. Natl. Acad. Sci. USA, 1990. 87: 1129-33; Akerlund et al., Cell division in Escherichia coli miniB mutants. Mol Microbiol., 1992, 6: 2073-2083).

[0244] Prokaryotes that have been shown to produce minicells include species of Escherichia, Shigella, Bacillus, Lac tobacillus, and Campylobacter. Bacterial minicell-producing species of particular interest are E. coli and Bacillus subtilis. E. coli is amenable to manipulation by a variety of molecular genetic methods, with a variety of well-characterized expression systems, including many episomal expression systems, factors and elements useful in the present invention. B. subtilis, also amenable to genetic manipulation using episomal expression elements, is an important industrial organism involved in the production of many of the world’s industrial enzymes (proteases, amylases, etc.), which it efficiently produces and secretes.

[0245] In the case of other eubacterial species, homologs of E. coli or B. subtilis genes that cause minicell production therein are known or can be identified and characterized as is known in the art. For example, the min regions of the chromosomes of Strepococcus pneumoniae and Neisseria gonorrhoeae have been characterized (Massidda et al., Unconventional organization of the division and cell wall gene cluster of Streptococcus pneumoniae, Microbiology 144:3069-78, 1998; and Ramirez-Arcos et al., Microbiology 147:225-237, 2001 and Szeto et al., Journal of Bacteria 183(21):6253, 2001, respectively). Those skilled in the art are able to isolate minicell producing (min) mutants, or prepare compounds inhibitory to genes that induce a minicell production (e.g., antisense to min transcripts).

TABLE 2

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Notes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campylobacter jejuni</td>
<td></td>
<td>may occur naturally late</td>
<td>Breck et al., 1987</td>
</tr>
<tr>
<td>B. subtilis</td>
<td></td>
<td>in growth cycle</td>
<td>Barak et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mutations in divIVB locus (inc. minC, minD)</td>
<td>Sciocetti et al., 1999; Lemon et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ripX mutations</td>
<td>Lemon et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>minC mutations</td>
<td>Moriya et al., 1998; Britton et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td>oriC deletions</td>
<td>Moriya et al., 1997; Hassan et al., 1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prfA mutations</td>
<td>Pederson and Seflow, 2001</td>
</tr>
<tr>
<td>Bacillus cereus</td>
<td>B1, 168</td>
<td>Mutations in divIVA locus</td>
<td>Chua et al., 1997; Sargent, 1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to initiation mutation Tsp143</td>
<td>Maier et al., 1999</td>
</tr>
<tr>
<td>Shigella flexneri</td>
<td>WC-1</td>
<td>Induced by exposure to long-chain polyphosphate</td>
<td>Genski et al., 1980</td>
</tr>
<tr>
<td>S. dysenteriae (2a)</td>
<td>MC-V</td>
<td></td>
<td>Genski et al., 1980</td>
</tr>
</tbody>
</table>

May 31, 2012
<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Notes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactobacillus spp.</td>
<td>Variant minicell-producing strains isolated from grains</td>
<td>Pidoux et al., 1990</td>
<td></td>
</tr>
<tr>
<td>Neisseria gonorrhoeae</td>
<td>deletion or overexpression of min homologues</td>
<td>Ramirez-Arcos et al., 2001; Szeto et al., 2001</td>
<td></td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>MinA mutations</td>
<td>Frazier et al., 1975; Cohen et al., 1976</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MinB mutations and deletions</td>
<td>Adler et al., 1967; Davies et al., 1984; Schumann et al., 1983; Jaffe et al., 1988; Akberud et al., 1992</td>
<td></td>
</tr>
<tr>
<td>CA800</td>
<td>cya, crp mutations</td>
<td>Kumar et al., 1979</td>
<td></td>
</tr>
<tr>
<td>MutM-1 mutation</td>
<td>MukE, mukF mutations</td>
<td>Yamanaka et al., 1996</td>
<td></td>
</tr>
<tr>
<td>LSU1 mutation</td>
<td>hns mutation</td>
<td>Kaidow et al., 1995</td>
<td></td>
</tr>
<tr>
<td>DS410</td>
<td>yf1972, yf1776 and yf2076</td>
<td>Heighway et al., 1989</td>
<td></td>
</tr>
<tr>
<td>P976-54</td>
<td>Temperature-sensitive cell division mutations</td>
<td>Curtiss, 1980</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by overexpression of minB protein</td>
<td>Adler et al., 1967; Allen et al., 1972; Hollesnberg et al., 1976; De Boer et al., 1988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by overexpression of minE protein or derivatives</td>
<td>Pichoff et al., 1995</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by overproduction of fnZ gene</td>
<td>Ward et al., 1985</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by overexpression of adaA gene</td>
<td>Wang et al., 1991</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by overexpression of min genes from <em>Neisseria gonorrhoeae</em></td>
<td>Ramirez-Arcos et al., 2001; Szeto et al., 2001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by exposure to EGTA</td>
<td>Wachi et al., 1999</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Induced by exposure to ampicillin</td>
<td>Elliot et al., 1985</td>
<td></td>
</tr>
</tbody>
</table>

Citations for Table 2:
Britten et al., Genes Dev. 12: 1254-9 (1998)
Cohen et al., Genetics 56: 550-551 (1967)
Frazier et al., Curr. Top. Immunol. 41: 81-84 (1975)
Hollesnberg et al., Gene 1: 23-47 (1997)
Moritz et al., DNA Res. 4: 115-26 (1997)
Wachi et al., Biochimie 81: 909-913 (1999)
Wang et al., Cell 42: 941-949 (1985)
II.A. Optimized Minicell Construction

Minicells are produced by several different eubacterial strains and mechanisms including the overexpression of endogenous or exogenous genes involved in cell division, chromosomal replication and partitioning, mutations in such genes, and exposure to various chemical and/or physical conditions. For example, in E. coli cells that overexpress the gene product FtsZ (the FtsZ gene encodes a protein that is involved in regulation of cell division; see Cook and Rothfield, Early stages in development of the Escherichia coli cell-division site. Mol Microbiol, 1994. 14: p. 485-495; and Lutkenhaus, Regulation of cell division in E. coli. Trends Genet, 1990. 6: p. 22-25), there is an increase in the formation of minicells (Hegg et al., Roles of FtsA and FtsZ in the activation of division sites. J Bacteriology, 1997. 180: 881-884). Minicells are also produced by E. coli cells having a mutation in one or more genes of the min locus, which is a group of genes that encode proteins that are involved in cell division (de Boer et al., Central role for the Escherichia coli minC gene product in two different cell division-inhibition systems. Proc Natl Acad Sci USA, 1990. 87: 1129-33; Akerlund et al., Cell division in Escherichia coli minB mutants. Mol Microbiol, 1992. 6: 2073-2083).

Eubacterial cells that have been shown to produce minicells include, but are not limited to species of Escherichia, Shigella, Bacillus, Lactobacillus, Legionella and Campylobacter. Bacterial minicell-producing species of particular interest are E. coli and Bacillus subtilis. These organisms are amenable to manipulation by a variety of molecular and genetic methods, with a variety of well-characterized expression systems, including many episomal and chromosomal expression systems, as well as other factors and elements useful in the present invention.

The following sections describe genes that may be manipulated so as to stimulate the production of minicells. The invention may include any of these non-limiting examples for the purpose of preparing minicells. Furthermore, these genes and gene products and conditions, may be used in methodologies to identify other gene(s), gene products, biological events, biochemical events, or physiological events that induce or promote the formation of minicells. These methodologies include, but are not limited to genetic selection, protein, nucleic acid, or combinatorial chemical libraries, screen, one- or two-hybrid analysis, display selection technologies, e.g. phage or yeast display, hybridization approaches, e.g. array technology, and other high- or low-throughput approaches.

II.A.1. Homologs

Homologs of these genes and gene products from other organisms may also be used. As used herein, a "homolog" is defined as a nucleic acid or protein having a nucleotide sequence or amino acid sequence, respectively, that is "identical," "essentially identical," "substantially identical," "homologous" or "similar" (as described below) to a reference sequence which may, by way of non-limiting example, be the sequence of an isolated nucleic acid or protein, or a consensus sequence derived by comparison of two or more related nucleic acids or proteins, or a group of isoforms of a given nucleic acid or protein. Non-limiting examples of types of isoforms include isoforms of differing molecular weight that result from, e.g., alternate RNA splicing or proteolytic cleavage; and isoforms having different post-translational modifications, such as glycosylation; and the like.

Two sequences are said to be "identical" if the two sequences, when aligned with each other, are exactly the same with no gaps, substitutions, insertions or deletions.

Two sequences are said to be "essentially identical" if the following criteria are met. Two amino acid sequences are "essentially identical" if the two sequences, when aligned with each other, are exactly the same with no gaps, insertions or deletions, and the sequences have only conservative amino acid substitutions. Conservative amino acid substitutions are as described in Table 3.

<table>
<thead>
<tr>
<th>Type of Amino Acid</th>
<th>Groups of Amino Acids that Are Conservative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short side chain</td>
<td>Hydrophobic, Lysine, Histidine, and tryptophan and tyrosine</td>
</tr>
<tr>
<td>Polar</td>
<td>Glutamine and Asparagine</td>
</tr>
<tr>
<td>Acidic</td>
<td>Glutamic Acid and Aspartic Acid</td>
</tr>
<tr>
<td>Basic</td>
<td>Arginine, Lysine and Histidine</td>
</tr>
</tbody>
</table>

Two nucleotide sequences are "essentially identical" if they encode the identical or essentially identical amino acid sequence. As is known in the art, due to the nature of the genetic code, some amino acids are encoded by several different three base codons, and these codons may thus be substituted for each other without altering the amino acid at that position in an amino acid sequence. In the genetic code, TTA, TTG, CTG, CTI, CTA and CGT encode Leu; AGA, AGG, CGT, CGC, CCA and CGG code Arg; GCT, GCC, GCA and GCG encode Ala; GTT, GCA, GGA and GGG encode Gly; ACT, ACC, ACA and ACG code Thr; GGT, GTC, GTA and GTG encode Val; TCT, TCC, TCA and TCG encode Ser; CCT, CCC, CCA and CCG encode Pro; ATA, ATC and ATA encode Ile; GAA and GAG encode Glu; CAA and CAG encode Gln; GAT and GAC encode Asp; AAT and AAC encode Asn; AGT and AGC encode Ser; TAT and TAC encode Tyr; TGTT and TGG encode Cys; AAA and AAG encode Lys; CAT and CAC encode His; TTT and TTC encode Phe; TGG encodes Trp; ATG encodes Met; and TGA, TAA and TAG are translation stop codons.

Two amino acid sequences are "substantially identical" if, when aligned, the two sequences are, (i) less than 30%, preferably ≤20%, more preferably ≤15%, most preferably ≤10%, of the identities of the amino acid residues vary between the two sequences; (ii) the number of gaps between or insertions in, deletions of and/or substitutions of, is ≥10%, more preferably ≤5%, more preferably ≤3%, most preferably ≤1%, of the number of amino acid residues that occur over the length of the shortest of two aligned sequences.

Two sequences are said to be "homologous" if any of the following criteria are met. The term "homolog" includes without limitation orthologs (homologs having genetic similarity as the result of sharing a common ancestor and encoding proteins that have the same function in different species) and parolog (similar to orthologs, yet gene and protein similarity is the result of a gene duplication).

One indication that nucleotide sequences are homologous is if two nucleic acid molecules hybridize to each other under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to
be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 M at pH 7 and the temperature is at least about 60°C.

Another way by which it can be determined if two sequences are homologous is by using an appropriate algorithm to determine if the above-described criteria for substantially identical sequences are met. Sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by algorithms such as, for example, the local homology algorithm of Smith and Waterman (Adv. Appl. Math. 2:482-489, 1981); by the homology alignment algorithm of Needleman and Wunsch (J. Mol. Biol. 48:443, 1970); by the search for similarity method of Pearson and Lipman (Proc. Natl. Acad. Sci. U.S.A. 85:2444, 1988); and by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, version 10.2 Genetics Computer Group (GCG), 575 Science Dr., Madison, Wis.); BLASTP, BLASTN, and FASTA (Altschul et al., J. Mol. Biol. 215:403-410, 1990); or by visual inspection.

Optimal alignments are found by inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482-489. "Gap" uses the algorithm of Needleman and Wunsch (1970 J Mol. Biol. 48:443-453) to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. In such algorithms, a "penalty" of about 3.0 to about 20 for each gap, and no penalty for end gaps, is used.

Homologous proteins also include members of clusters of orthologous groups of proteins (COGs), which are generated by phylogenetic classification of protein sequences encoded in complete genomes. To date, COGs have been delineated by comparing protein sequences encoded in 43 complete genomes, representing 30 major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain (see Tatusov et al., A genomic perspective on protein families. Science. 278: 631-637, 1997; Tatusov et al., The COG database: a tool for genome-scale analysis of protein functions, and gene families. Nucleic Acids Res. 29:22-28, 2001; Chervitz et al., Comparison of the Complete Sets of Worm and Yeast: Orthology and Divergence. Science 282:2022-2028, 1998; and http://www.ncbi.nlm.nih.gov/COG/).

The entirety of two sequences may be identical, essentially identical, substantially identical, or homologous to one another, or portions of such sequences may be identical or substantially identical with sequences of similar length in other sequences. In either case, such sequences are similar to each other. Typically, stretches of identical or essentially within similar sequences have a length of \( \geq 12 \), preferably \( \geq 24 \), more preferably \( \geq 48 \), and most preferably \( \geq 96 \) residues.

II.A.2. Escherichia coli Genes


0270) II.B. Gene Expression in Minicells

0271) II.B.1. In General

0272) In some aspects of the invention, it may be desirable to alter the expression of a gene and the production of the corresponding gene product. As is known in the art, and is used herein, a "gene product" may be a protein (polypeptide) or nucleic acid. Genes products that are proteins include without limitation enzymes, receptors, transcription factors, termination factors, expression factors, DNA-binding proteins, proteins that effect nucleic acid structure, or subunits of any of the preceding. Gene products that are nucleic acids include, but are not limited to, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), antisense RNAs, nucleases (including but not limited to catalytic RNAs, ribonucleases, and the like).

0273) Depending on the function of a gene product, and on the type of application of the invention, it may be desirable to increase protein production, decrease protein production, increase protein nucleic acid production and/or increase nucleic acid production. Provided herein are non-limiting examples of genes and gene products that may be manipulated, individually or in combination, to modulate the expression of gene products to be included into minicells or parent strains from which minicells are derived. The expression elements so modulated may be chromosomal and/or episomal, and may be expressed constitutively or in a regulated fashion, i.e., repressible and/or inducible. Furthermore, gene products under the regulation may be either monocistronic or polycistronic with other genes or with themselves.

0274) II.B.2. Protein Production

0275) By way of non-limiting example, increased protein production may occur through increased gene dosage (increased copy number of a given gene under the control of the native or artificial promotor where this gene may be on a plasmid or in more than one copy on the chromosome), modification of the native regulatory elements, including, but not limited to the promotor or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/enhancers, or relevant antisense nucleic acid or nucleic acid analog, cloning on a plasmid under the control of the native or artificial promotor, and increased or decreased production of native or artificial promotor regulatory element(s) controlling production of the gene or gene product.

0276) By way of non-limiting example, decreased protein production may occur through modification of the native regulatory elements, including, but not limited to the promotor or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/enhancers, or relevant antisense nucleic acid or nucleic acid analog, through cloning on a plasmid under the control of the native regulatory region containing mutations or an artificial promotor, either or both of which resulting in decreased protein production, and through increased or decreased produc-
tion of native or artificial promoter regulatory element(s) controlling production of the gene or gene product.

[0282] As used herein with regards to nucleic acids, “intramolecular activity” refers to a structure-dependent function. By way of non-limiting example, alteration of intramolecular activity may be accomplished by mutation of the gene, in vivo or in vitro chemical modification of the nucleic acid, inhibitor molecules against the function of the nucleic acid, e.g., competitive, non-competitive, or uncompetitive enzymatic inhibitors, inhibitors that prevent protein-nucleic acid interactions, e.g., expression or introduction of dominant-negative or dominant-positive protein or other nucleic acid fragment(s), or other carbohydrate(s), fatty acid(s), and lipid(s) that may act directly or allosterically upon the nucleic acid or nucleic acid-protein complex, and/or modification of nucleic acid moieties that modify the gene or gene product to create the functional nucleic acid.

[0283] As used herein with regards to nucleic acids, “intramolecular function” refers to the effects resulting from an intramolecular interaction between the nucleic acid and another nucleic acid, protein, carbohydrate, fatty acid, lipid, or other molecule(s) in or on the cell or the action of a product or products resulting from such an interaction. By way of non-limiting example, intramolecular function may be the act or result of intramolecular or intramolecular phosphorylation, biotinylation, methylation, acetylation, glycosylation, and/or other signaling event; this function may be the result of a protein-nucleic acid, and/or carrier function, e.g., the capacity to bind, either covalently or non-covalently small organic or inorganic molecules, protein(s), carbohydrate(s), fatty acid(s), lipid(s), and other nucleic acid(s); this function may be to interact with the membrane to recruit other molecules to this compartment of the cell; this function may be to regulate the transcription and/or translation of the gene, other nucleic acid, or protein; and this function may be to stimulate the function of another process that is not yet described or understood.

[0284] II.C. Genes and Gene Products for Regulation of Expression

[0285] As is known in the art, a variety of genes, gene products and expression elements may be manipulated, individually or in combination, in order to modulate the expression of genes and/or production gene products. These include, by way of non-limiting example, RNA polymerases, ribosomes (ribosomal proteins and ribosomal RNAs), transfer RNAs (tRNAs), amino transferases, regulatory elements and promoter regions, transportation of inducible and inhibitory compounds, catabolite repression, general deletions and modifications, cytoplasmic redox state, transcriptional terminators, mechanisms for ribosomal targeting, proteins, chaperones, export apparatus and membrane targeting, and mechanisms for increasing stability and solubility. Each of these is discussed in more detail in the following sections.

[0286] II.C.I. RNA Polymerases

[0287] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include modification of an endogenous and/or introduction of an exogenous RNA polymerase. A rho gene, or any other gene that encodes a RNA polymerase subunit product from E. coli, or homologs of this gene or its gene product found in other prokaryotes, eukaryotes, archaeabacteria or organelles (mitochondria, chloroplasts, plastids and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells.

[0288] The production or activity of a desired gene product may be increased by increasing the level and/or activity of an RNA polymerase that transcribes the gene product’s cognate gene. The production or activity of a desired protein gene product may be increased by decreasing the level and/or activity of an RNA polymerase that transcribes a gene product that inhibits the production or function of the desired gene product.

[0289] As one example, manipulation of the rpoA (plus, see) gene or gene product from E. coli, or homologs of this gene or gene product found in other members of the Prokaryotes, Eukaryotes, Archaeabacteria and/or organelles (e.g., mitochondria, chloroplasts, plastids and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells. In addition to rpoA, E. coli genes that encode RNA polymerase subunits include rpoB (fisR, groN, nitB, rif, ron, stl, sty, tabD, sgdB, mbrD), rpoC (tabD), rpoD (alt), rpoE, rpoH (fam, hin, htpR), rpoN (ginF, ntrA), rpoS (abrD, dpeB, katF, nur), and rpoZ (spoS). See Berlyn et al., “Linkage Map of Escherichia coli K-12, Edition 9,” Chapter 109 in: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 2, pages 1715-1902, and references cited therein; and Sanderson et al., “Linkage Map of Salmonella typhimurium, Edition VIII” Chapter 110 in: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 2, pages 1903-1999, and references cited therein.

[0290] Production of a desired gene product may be preferentially or selectively enhanced by the introduction of an exogenous RNA polymerase that specifically recognizes expression sequences that are operably linked to the corresponding gene. By way of non-limiting example, the use of a T7 RNA polymerase to selectively express genes present on expression elements that segregate into minicells is described herein.

[0291] II.C.2. Ribosomes

[0292] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include modification of endogenous, and/or addition of exogenous, ribosomes or ribosomal subunits. The techniques may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells.

[0293] As is known in the art, a ribosome includes both proteins (polypeptides) and RNA (rRNA). Thus, in the case of a gene that encodes a component of a ribosome, the gene product may be a protein or an RNA. For a review, see Noller et al., “Ribosomes,” Chapter 13 in: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 1, pages 167-186, and references cited therein. For the sake of convenience, both ribosomal proteins and rRNAs are encompassed by the term “ribosomal subunits.”
The production or activity of a desired protein gene product may be increased by increasing the level and/or activity of a ribosomal subunit that causes or enhances the translation of the desired protein. The production or activity of a desired protein gene product may be increased by decreasing the level and/or activity of a ribosomal subunit that causes or enhances translation of a protein that has a negative impact on the production or activity of the desired protein.


Transfer RNAs (tRNAs)

In the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of endogenous and/or exogenous transfer RNAs (tRNAs). Manipulation of the tRNA genes or gene products from E. coli, or homologs of tRNA genes or gene products found in other members of the Prokaryotes, Eukaryotes, Archaeabacteria and organelles (including but not limited to mitochondria, chloroplasts, plastids, and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or segregated minicells.

Exemplary E. coli tRNA genes include, but are not limited to, the talaF (talaA) gene, the talaU (talaD) gene, the talaV (talaB) gene, the talaW (talaC) gene, the talaX (talaJ) gene, the talaY (talaK) gene, the talaZ (talaE) gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene, the talaA gene, the talaB gene, the talaC gene, the talaD gene, the talaE gene, the talaF gene, the talaG gene, the talaH gene, the talaI gene, the talaJ gene, the talaK gene, the talaL gene, the talaM gene, the talaN gene, the talaO gene, the talaP gene, the talaQ gene, the talaR gene, the talaS gene, the talaT gene, the talaU gene, the talaV gene, the talaW gene, the talaX gene, the talaY gene, the talaZ gene.
protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of endogenous and/or exogenous aminoacyl synthetases and proteins that effect their production and/or activity. Aminoacyl synthetases are involved in "charging" a tRNA molecule, i.e., attaching a tRNA to its cognate amino acid. (Martinis et al., Aminoacyl-tRNA Synthetases: General Features and Relationships. Chapter 58 in: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology; 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 1, pages 887-901) and references cited therein; (Grunberg-Manago, Regulation of the Expression of Aminoacyl-tRNA Synthetases and Translation. Chapter 91 in: Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology; 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 1, pages 1432-1457) and references cited therein; and (Hershey, "Protein Synthesis," Chapter 40 in: Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology, Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1987, Volume 1, pages 613-647) and references cited therein.

**[0304]** By way of non-limiting example, manipulation of the aat gene or gene product from E. coli, or homologs of this gene or gene product found in other members of the Prokaryotes, Eukaryotes, Archaea bacteria and/or organelles (e.g., mitochondria, chloroplasts, plastids and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells (Boehner, B. R., and Savageau, M. A. 1979. Inhibition of growth by imidazolone propionic acid: evidence in vivo for coordination of histidine catabolism with the catabolism of the other amino acids. Mol. Gen. Genet. 168(1):87-95).


**[0306]** II.C.5. Regulatory Elements and Promoter Regions

**[0307]** Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of regulatory elements and promoter regions. Such manipulations may result in increased or decreased production, and/or changes in the intramolecular and intermolecular functions, of a protein in a segregated minicell or its parent cell prior to minicell formation; in the latter instance, the protein may be one that is desirable retained in segregated minicells.

**[0308]** The production or activity of a desired gene product may be increased by increasing the level and/or activity of a promoter or other regulatory region that acts to stimulate or enhance the production of the desired gene product. The
production or activity of a desired gene product may be increased by decreasing the level and/or activity of a promoter or other regulatory region that acts to stimulate or enhance the production of a gene product that acts to reduce or eliminate the level and/or activity of the desired gene product.

[0309] 11. C.S.a. Escherichia coli


May 31, 2012

**[Bacillus subtilis]**

**Regulatory elements, promoters and other expression elements from B subtilis**


[0314] II.C.5.e. Other Eubacteria


[0318] II.C.5.e. Use of Site-Specific Recombination in Expression Systems

[0319] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting examples, these techniques may include modification of endogenous and/or exogenous regulatory elements responsible for activation and/or repression of proteins to be expressed from chromosomal and/or plasmid expression vectors. By way of non-limiting example, this system may be applied to any of the above regulatory elements/systems. Specifically, each of the above mentioned regulatory systems may be constructed such that the promoter regions are oriented in a direction away from the gene to be expressed, or each of the above mentioned gene(s) to be expressed may be constructed such that the gene(s) to be expressed is oriented in a direction away from the regulatory region promoter. Constructed in this system is a methodology dependent upon site-specific genetic recombination for inversion and induction of the gene of interest (Backman, K., et al. 1984. Use of synchronous site-specific recombination in vivo to regulate gene expression. Bio/Technology 2:1045-1049; Balakrishnan, R., et al. 1994. A gene cassette for adapting Escherichia coli strains as hosts for att-Int-mediated rearrangement and pl. expression vectors. Gene 138:101-104; Hasan, N., and W. Szybalaki. 1987. Control of cloned gene expression by promoter inversion in vivo: construction of improved vectors with a multiple cloning site and the Puc promoter. Gene 56:145-151; Wulfing, C., and A. Pluckthun. 1993. A versatile and highly repressible Escherichia coli expression system based on invertible promoters: expression of a gene encoding a toxic gene product. Gene 136:199-203). These invertible promoters and/or gene regions will allow tight regulation of potentially toxic protein products. By way of non-limiting example, these systems may be derived from bacteriophage lambda, bacteriophage Mu, and/or bacteriophage P22. In any of these potential systems, regulation of the recombination may be regulated by any of the regulatory systems discussed in section II.C.5 and elsewhere herein.

[0320] II.L.C.5.e. Use of Copy Number Control Switches

[0321] A method that can be used to increase the efficiency of gene expression and protein production in minicells involves the modification of endogenous and/or introduction of exogenous genetic expression systems such that the number of copies of a gene encoding a protein to be expressed can be modulated. Copy number control systems comprise elements designed to modulate copy number in a controlled fashion.

[0322] In an exemplary mode, copy number is controlled to decrease the effects of “leaky” (uninduced) expression of toxic gene products. This allows one to maintain the integrity of a potentially toxic gene product during processes such as cloning, culture maintenance, and periods of growth prior to minicell-induction. That is, decreasing the copy number of a gene is expected to decrease the opportunity for mutations affecting protein expression and/or function to arise. Immediately prior to, during and/or after minicell formation, the copy number may be increased to optimize the gene dosage in minicells as desired.


[0324] By way of non-limiting example, the pcnB gene product, the wildtype form of which promotes increased ColE1 plasmid copy number (Soderbon, F., et al. 1997. Regulation of plasmid R1 replication: PcnB and Rnase E expedite the decay of the antisense RNA, CopA. Mol. Microbiol. 26:493-504), is modulated; and/or mutant forms of the pcnB gene are introduced into a cell. In an exemplary cell type that may be used in the methods of the invention, the wildtype pcnB chromosomal gene is replaced with a mutant pcnB80 allele (Lopilato, J., et al. 1986. Mutations in a new chromosomal gene of Escherichia coli K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives. Mol. Gen. Genet. 205:285-290). In such cells the copy number of a ColE1-derived plasmid is decreased. The cell may further comprise an expression element comprising an inducible promoter operably linked to an ORF encoding the wild-type pcnB. Because the wild-type pcnB gene is dominant to the mutant pcnB80 gene, and because the wild-type pcnB gene product promotes increased ColE1 plasmid copy number, induction of a wild-type pcnB in the pcnB80 background will increase the plasmid copy number of ColE1-derived plasmids. Such copy number control systems may be expressed from the chromosome and/or plasmid to maintain either low or high
plasmid copy number in the absence of induction. Other non-limiting examples of gene and/or gene products that may be employed in copy number control systems for ColE1-based replicons include genes or homologs of genes encoding RNA I, RNA II, rpp, RNase H, enzymes involved in the process of polyadenylation, RNase E, DNA polymerase I, and DNA polymerase III.

[0325] In the case of IncFII-derived replicons, non-limiting examples of gene and/or gene products that may be employed in copy number control systems to control plasmid copy include genes or homologs of the copA, copB, repA, and repB genes. Copy number control systems may additionally or alternatively include manipulation of repC, trpA, dnaA, dnaB, dnaC, seqA, genes protein Pi, genes encoding HU protein subunits (hupA, hupB) and genes encoding IHF subunits.

[0326] Other elements may also be included to optimize these plasmid copy number control systems. Such additional elements may include the addition or deletion of iteron nucleic acid sequences (Chattonoy, D. K. 2000. Control of plasmid DNA replication by iterons: no longer paradoxic. Mol. Microbiol. 37:467-476), and modification of chaperone proteins involved in plasmid replication (Konieczny, J., et al. 1997. The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpXP chaperone. Proc Natl Acad Sci USA 94:14378-14382).

[0327] II.C.6. Induction of Inhibitory Compounds

[0328] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of factors and systems that modulate the transport of compounds, including but not limited to inducers and/or inhibitors of expression elements that control expression of a gene in a parent cell prior to minicell formation and/or in segregated minicells. Such manipulations may result in increased or decreased production, and/or changes in the intramolecular and intermolecular functions, of a protein in a minicell or its parent cell. The techniques may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells.

[0329] II.C.6.a. Escherichia coli Genes

[0330] By way of non-limiting example, manipulation of the abpS gene or gene product from E. coli, or homologs of this gene or gene product found in other members of the Prokaryotes, Fakeryotes, Archaeabacteria and/or organelles (e.g., mitochondria, chloroplasts, plastids and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells (Celis, R. T. 1982. Mapping of two loci affecting the synthesis and structure of a periplasmic protein involved in arginine and ornithine transport in Escherichia coli K-12. J. Bacteriol. 151(3):1314-9).


**[0335]** II.C.7. Catabolite Repression

**[0336]** Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of factors and systems involved in the synthesis, degradation or transport of catabolites that modulate the genetic expression of a preselected protein. Such manipulations may result in increased or decreased production, and/or changes in the intramolecular and intermolecular functions, of a protein in a minicell or its parent cell; in the latter instance, the protein may be one that is desirably retained in segregated minicells.

**[0337]** By way of non-limiting example, it is known in the art to use promoters from the trp, cst-1, and lbp operons of *E. coli*, which are induced by, respectively, reduced tryptophan levels, glucose starvation, and lactose. Manipulation of the catabolites tryptophan, glucose and lactose, respectively, will influence the degree of expression of genes openably linked to these promoters. (Makrides, Sarvas C., Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*. Microbiological Reviews. 1996. 60:512-538.)

**[0338]** As another non-limiting example, expression elements from the *E. coli* L-arabinose (ara) operon are used in expression systems. AraC is a protein that acts as a repressor of ara genes in the absence of arabinose, and also as an activator of ara genes when arabinose is present. Induction of ara genes also involves cAMP, which modulates the activity of CRP (cAMP receptor protein), which in turn is required for full induction of ara genes (Schleif, Robert, Regulation of the L-arabinose operon of *Escherichia coli*. 2000. TIG 16:559-564. Thus, maximum expression from an ara-based expression system is achieved by adding cAMP and arabinose to host cells, and optimizing the expression of CRP in hostcells. (0339) As one example, manipulation of the acpS gene or gene product of *E. coli* (Pollacco M. L., and J. E. Cronan Jr. 1981. A mutant of *Escherichia coli* conditionally defective in the synthesis of holo-[acyl carrier protein]. J. Biol. Chem. 256:5750-5754; or homologs of this gene or its gene product found in other prokaryotes, such as *Bacillus* species. Deletion of organelles (mitochondria, chloroplasts, plastids and the like) may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicell formation and/or in segregated minicells.

occur through increased gene dosage (increased copy number of a given gene under the control of the native or artificial promoter where this gene may be on a plasmid or in more than one copy on the chromosome), modification of the native regulatory elements, including, but not limited to the promoter or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/inhibitors, or relevant antisense nucleic acid or nucleic acid analog, cloning on a plasmid under the control of the native or artificial promoter; and increased or decreased production of native or artificial promoter regulatory elements controlling production of the gene. By way of non-limiting example, decreased gene expression production may occur through modification of the native regulatory elements, including, but not limited to the promoter or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/inhibitors, or relevant antisense nucleic acid or nucleic acid analog, through cloning on a plasmid under the control of the native regulatory region containing mutations or an artificial promoter, either or both of which resulting in decrease gene expression, and through increased or decreased production of native or artificial promoter regulatory element(s) controlling gene expression. By definition, intramolecular activity refers to the enzymatic function, structure-dependent function, e.g. the capacity of a gene product to interact in a protein-protein, protein-nucleic acid, or protein-lipid complex, and/or carrier function, e.g. the capacity to bind, either covalently or non-covalently small organic or inorganic molecules, protein(s) carbohydrate(s), fatty acid(s), lipid(s), and nucleic acid(s). By way of non-limiting example, alteration of intramolecular activity may be accomplished by mutation of the gene, in vivo or in vitro chemical modification of the gene product, inhibitor molecules against the function of the gene product, e.g. competitive, non-competitive, or uncompetitive enzymatic inhibitors, inhibitors that prevent protein-protein, protein-nucleic acid, or protein-lipid interactions, e.g. expression or introduction of dominant-negative or dominant-positive or other protein fragment(s), or other carbohydrate(s), fatty acid(s), lipid(s), and nucleic acid(s) that may act directly or allosterically upon the gene product, and/or modification of protein, carbohydrate, fatty acid, lipid, or nucleic acid moieties that modify the gene or gene product to create the functional protein. By definition, physiological function refers to the effects resulting from an intramolecular interaction between the gene product and other protein, carbohydrate, fatty acid, lipid, nucleic acid, or other molecule(s) in or on the cell or the action of a product or products resulting from such an interaction.

By way of non-limiting example, physiological function may be the act or result of intermolecular phosphorylation, biotinilation, methylation, acylation, glycosylation, and/or other signaling event; this function may be the result of protein-protein, protein-nucleic acid, or protein-lipid interaction resulting in a functional moiety; this function may be to interact with the membrane to recruit other molecules to this compartment of the cell; this function may be to regulate the transcription and/or translation of trxA, another protein, or nucleic acid; and this function may be to stimulate the function of another process that is not yet described or understood.

In the design of the invention are techniques that increase the efficiency of gene expression and protein production in parental cell cytoplasm prior to minicell formation and/or the segregated minicell cytoplasm. By way of non-limiting example, these techniques may include modification of terminator regions of DNA templates or RNA transcripts so that transcription and/or translation of these nucleic acid regions will terminate at greater efficiency. By way of non-limiting example, these techniques may include stem-loop structures, consecutive translational terminators, polyadenylation sequences, or increasing the efficiency of ribo-dependent termination. Stem loop structures may include, but are not limited to, inverted repeats containing any combination of deoxyribonucleic acid or ribonucleic acid molecule, more than one such inverted repeat, or variable inverted repeats such that the rate of transcriptional/translational termination may be moderated dependent on nucleic acid and/or amino acid concentration, e.g. the mechanism of regulatory attenuation (Oxender et al., Attenuation in the Escherichia coli tryptophan operon: role of RNA secondary structure involving the tryptophan codon region, Proc. Natl. Acad. Sci. 76:5524-5528, 1979). See also, Yager and von Hippel, “Transcription Elongation and Termination in e. Col. And Landick and Yanofsky, “Transcriptional Attenuation,” Chapters 76 and 77, respectively in: Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology, Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1987, Volume 1, pages 1241-1275 and 1276-1301, respectively, and references cited therein.

In the design of the invention are techniques that increase the efficiency of gene expression and protein production in parental cell cytoplasm prior to minicell formation and/or the segregated minicell cytoplasm. By way of non-limiting example, these techniques may include modifications of endogenous and/or exogenous ribosomal components such that ribosomes enter the minicell segregates with higher efficiency. By way of non-limiting example, these techniques may include increasing the copy number of ribosomal binding sites on plasmid or like structure to recruit more ribosomal components or increase the synthesis of ribosomal subunits prior to segregation (Mawn et al., Depletion of free 50S ribosomal subunits in Escherichia coli by expression of RNA containing Shine-Dalgarno-like sequences, J. Bacteriol. 184:494-502, 2002). This construct may also include the use of plasmid expressed translation initiation factors to assist ribosomal segregation (Celano et al., Interaction of Escherichia coli translation-initiation factor IF-1 with ribosomes, Eur. J. Biochem. 178:351-355 1988). See also Hoopes and McClure, “Strategies in Regulation of Transcription Initiation,” Chapter 75 in: Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology, Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1987, Volume 2, pages 1231-1240, and references cited therein.

In the design of the invention are techniques that increase the efficiency of gene expression and protein production in minicells. By way of non-limiting example, these techniques may include utilization and/or modification of endogenous and/or exogenous proteases. Such manipulations may result in increased or decreased production, and/or changes in the intramolecular and intermolecular functions, of a protein in a minicell or its parent cell; in the latter instance, the protein may be one that is desirably retained in segregated minicells.
The production or activity of a desired protein gene product may be increased by decreasing the level and/or activity of a protease that acts upon the desired protein. The production or activity of a desired protein gene product may be increased by increasing the level and/or activity of a protease that acts upon a protein that inhibits the production or function of the desired protein.

The production or activity of a desired nucleic acid gene product may be increased by increasing the level and/or activity of a protease that acts upon a protein that inhibits the production or function of the nucleic acid gene product. The production or activity of a desired nucleic acid gene product may be increased by decreasing the level and/or activity of a protease that acts upon a protein that stimulates or enhances the production or function of the desired nucleic acid gene product.

As one example, manipulation of the αlP gene or gene product from E. coli (Kirby J. E., and J. E. Trempey, and S. Gottesman. Excision of a P4-like cryptic prophage leads to Alp protease expression in Escherichia coli. 1994. J Bacteriol. 176:2068-2081), or homologs of this gene or gene product found in other members of the Prokaryotes, Eukaryotes or Archaeabacteria, may be employed to increase the efficiency of gene expression and protein production in parent cells prior to minicellular formation and/or segregated minicellular postpartum.


[0356] II.C.13. Chaperones

[0357] Included in the design of the invention are techniques that increase the efficiency of gene expression and functional protein production in microcells. By way of non-limiting example, these techniques may include modification of chaperones and chaperonins, i.e., endogenous and/or exogenous protein components that monitor the unfolded state of translated proteins allowing proper folding and/or secretion, membrane insertion, or soluble multimeric assembly of expressed proteins in the parental cell prior to micelle formation and/or the segregated micelle cytoplasm, membrane, periplasm, and/or extracellular environment. See Gottesman et al., Protein folding and unfolding by *Escherichia coli* chaperones and chaperonins, Current Op. Microbiol. 3:197-202, 2000; and Mayhew et al., “Molecular Chaperone Proteins,” Chapter 61 in: *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd Ed., Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1996, Volume 1, pages 922-937, and references cited therein.

[0358] These applications may, but are not limited to increased or decreased chaperone production, increased or decreased intramolecular activity of a chaperone, increased or decreased physiological function of a chaperone, deletion, substitution, inversion, translocation or insertion into, or mutation of, a gene chaperone. By way of non-limiting example, increased production of a chaperone may occur through increased chaperone gene dosage (increased copy number of a given gene under the control of the native or artificial promoter where this gene may be on a plasmid or in more than one copy on the chromosome), modification of the native regulatory elements, including, but not limited to the promoter or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/enhancers, or relevant antisense nucleic acid or nucleic acid analog, cloning on a plasmid under the control of the native or artificial promoter, and increased or decreased production of native or artificial promoter regulatory element(s)
controlling production of the chaperone gene or gene product. By way of non-limiting example, decreased production of a chaperone may occur through modification of the native regulatory elements, including, but not limited to the promoter or operator region(s) of DNA, or ribosomal binding sites on RNA, relevant repressors/silencers, relevant activators/enhancers, or relevant antisense nucleic acid or nucleic acid analog, through cloning on a plasmid under the control of the native regulatory region containing mutations or an artificial promoter, either or both of which resulting in decreased chaperone production, and through increased or decreased production of native or artificial promoter regulatory element(s) controlling production of the chaperone gene or gene product. By definition, intramolecular activity refers to the enzymatic function, structure-dependent function, e.g. the capacity of chaperone to interact in a protein-protein, protein-nucleic acid, or protein-lipid complex, and/or carrier function, e.g. the capacity to bind, either covalently or non-covalently small organic or inorganic molecules, protein(s), carbohydrate(s), fatty acid(s), lipid(s), and nucleic acid(s). By way of non-limiting example, alteration of intramolecular activity may be accomplished by mutation of the chaperone gene, in vivo or in vitro chemical modification of Chaperone, inhibitor molecules against the function of chaperone, e.g. competitive, non-competitive, or competitive enzyme inhibitors, inhibitors that prevent protein-protein, protein-nucleic acid, or protein-lipid interactions, e.g. expression or introduction of dominant-negative or dominant-positive chaperone or other protein fragment(s), or other carbohydrate(s), fatty acid(s), lipid(s), and nucleic acid(s) that may act directly or allosterically upon Chaperone, and/or modification of protein, carbohydrate, fatty acid, lipid, or nucleic acid moieties that modify the chaperone gene or gene product to create the functional protein. By definition, physiological function refers to the effects resulting from an intramolecular interaction between Chaperone and other protein(s), carbohydrate, fatty acid, lipid, nucleic acid, or other molecule(s) in or on the cell or the action of a product or products resulting from such an interaction. By way of non-limiting example, physiological function may be the act or result of intramolecular phosphorylation, biotinylation, methylation, acylation, glycosylation, and/or other signaling event; this function may be the result of a protein-protein, protein-nucleic acid, or protein-lipid interaction resulting in a functional moiety; this function may be to interact with the membrane to recruit other molecules to this compartment of the cell; this function may be to regulate the transcription and/or translation of chaperone, other protein, or nucleic acid; and this function may be to stimulate the function of another process that is not yet described or understood.


[0360] II.C.14. Export Apparatus and Membrane Targeting

[0361] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in parental cells prior to minicell formation and/or in the segregated minicells. By way of non-limiting example, these techniques may include construction of chimeric proteins including, but not limited to, coupling the expressed protein of interest with native Eubacterial, Eukaryotic, Archeabacterial or organellar leader sequences to drive membrane insertion or secretion of the protein of interest to the periplasm or extracellular environment. In addition to using native leader sequences, these minicell expression constructs may also include proteolytic cleavage sites to remove the leader sequence following insertion into the membrane or secretion. These proteolytic cleavage sites may be native to the organism from which the minicell is derived or non-native. In the latter example, also included in the system are the non-native protease that recognizes the non-native proteolytic cleavage site.


[0363] In addition to these leader sequences, mutations in the cellular export machinery may be employed to increase the promiscuity of export to display or export sequences with non-optimized leader sequences. Non-limiting examples of genes that may be altered to increase export promiscuity are mutations in secY (pRlA4) (Derman, A. L., et al. 1993. A signal sequence is not required for protein export in plmA mutants of Escherichia coli. EMBO J. 12:879-888), and secE (Harris, C. R., and T. J. Silhavy. 1999. Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and in vivo cross-linking. J. Bacteriol. 181:3438-3444).

[0364] II.C.15. Increasing Stability and Solubility

[0365] Included in the design of the invention are techniques that increase the efficiency of gene expression and protein production in parental cells prior to minicell formation and/or in the segregated minicells. By way of non-limiting example, these techniques may include construction of chimeric/fusion proteins including, but not limited to, coupling the expressed protein of interest with native Eubacterial, Eukaryotic, Archeabacterial or organellar solubilizing sequences. As used herein, “solubilizing sequences” are complete or truncated amino acid sequences that increase the solubility of the expressed membrane protein of interest. This increased solubility may be used to increase the lifetime of the soluble state until proper membrane insertion may take place. By way of non-limiting example, these solubile chimeric fusion proteins may be ubiquitin (Power, R. F., et al. 1990. High level expression of a truncated chicken progesterone receptor in Escherichia coli. J. Biol. Chem. 265:1419-1424), thioredoxin (LaValle, E. R., et al. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology (N.Y.) 11:187-193; Kapust, R. B., and D. S. Waugh. 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8:1668-1674), the dsbA gene product (Winter, J., et al. 2001. Increased production of human proinsulin in the periplasmic space of Escherichia coli by fusion to DsbA. J. Biotechnol. 84:175-185), the SPG protein (Murphy, J. P., et al. 1992. Amplified expression and large-scale purification of protein G. Bioseparation 3:63-71), the mleE gene product (maltose-binding protein) (Hamp, W., et al. 2000. Engineering of a proteolytically stable human beta 2-adrenergic receptor/maltose-binding protein fusion and production of the chimeric protein in Escherichia coli and baculovirus-infected insect cells. J. Biotechnol. 77:219-234; Kapust et al., Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 8:1668-1674, 1999), glutathione-s-transferase (GST); and/or nuclease A (Mecker et al., A fusion protein, between serum amyloid A and staphylococcal nuclease—synthesis, purification, and structural studies, Proteins 30:381-387, 1998). In addition to these proteins, Staphylococcal protein A, beta-galactosidase, S-peptide, myosin heavy chain, dihydrofolate reductase, T4 p55, growth hormone N terminus, E. coli Hemolysin A, bacteriophage lambda cI protein, TrpL, and TrpLE proteins may also be used as fusion proteins to increase protein expression and/or solubility (Makrides, Strategies for Achieving High-Level Expression of Genes in Escherichia coli, Microbiol. Rev. 60:512-538).

[0366] III. Preparation of Minicells

[0367] III.A. Parent Cell Mutations

[0368] Although it has been reported that relatively few molecules of endogenous RNA polymerase segregate into minicells (Shepherd et al., Cytoplasmic RNA Polymerase in Escherichia coli, J. Bacteriol. 183:2527-34, 2001), other reports and results indicate that many RNA Polymerase molecules follow plasmids into minicells (Funnell and Gagnier, Partition of PI plasmids in Escherichia coli mukB chromosomal partition mutants, J. Bacteriol. 177:2381-6, 1995). In any event, applicants have discovered that the introduction of an exogenous RNA polymerase to minicell-producing cells enhances expression of episomal elements in minicells. Such enhanced expression may allow for the successful expression of proteins in minicells, wherein such proteins are expressed poorly or not at all in unmodified minicells. In order to maximize the amount of RNA transcription from episomal elements in minicells, minicell-producing cell lines that express an RNA polymerase specific for certain episomal expression elements may be used. An example of an E. coli strain of this type, designated MC-T7, was created and used as is described
in the Examples. Those skilled in the art will be able to make and use equivalent strains based on the present disclosure and their knowledge of the art.

[0369] Minicell-producing cells may comprise mutations that augment preparative steps. For example, lipopolysaccharide (LPS) synthesis in E. coli includes the lipid A biosynthetic pathway. Four of the genes in this pathway have now been identified and sequenced, and three of them are located in a complex operon that also contains genes involved in DNA and phospholipid synthesis. The rfa gene cluster, which contains many of the genes for LPS core synthesis, includes at least 17 genes. The rfb gene cluster encodes protein involved in O-antigen synthesis, and rfb genes have been sequenced from a number of serotypes and exhibit the genetic polymorphism anticipated on the basis of the chemical complexity of the O antigens. See Schnaitman and Klena, Genetics of lipopolysaccharide biosynthesis in enteric bacteria, Microbiol. Rev. 57:655-82, 1993. When present, alone or in combination, the rfb and oms mutations cause alterations in the eubacterial membrane that make it more sensitive to lysozyme and other agents used to process minicells. Similarly, the rfa (Chen, L., and W. G. Coleman Jr. 1993. Cloning and characterization of the Escherichia coli K-12 rfa-2 (rfaC) gene, a gene required for lipopolysaccharide inner core synthesis, J. Bacteriol. 175:2534-2540), lpcA (Brooke, J. S., and M. A. Valvano. 1996. Biosynthesis of inner core lipopolysaccharide in enteric bacteria identification and characterization of a conserved phosphoprotease isomerase, J. Biol. Chem. 271:3608-3614), and lpcB (Kadman, J. L., et al. 1998. Cloning and overexpression of glycosyltransferases that generate the lipopolysaccharide core of Rhizobium leguminosarum. J. Biol. Chem. 273:26432-26440) mutations, when present alone or in combination, cause alterations in lipopolysaccharides in the outer membrane causing cells to be more sensitive to lysozyme and agents used to process minicells. In addition, such mutations can be used to reduce the potential antigenicity and/or toxicity of minicells.

[0370] III.B. Culturing Conditions

[0371] Included in the design of the invention are the conditions to grow parental cells from which minicells will be produced. Non-limiting examples herein are drawn to conditions for growing E. coli parental cells to produce minicells derived from E. coli parental cells. Non-limiting examples for growth media may include rich media, e.g. Luria broth (LB), defined minimal media, e.g. M63 salts with defined carbon, nitrogen, phosphate, magnesium, and sulfate levels, and complex minimal media, e.g. defined minimal media with casamino acid supplement. This growth may be performed in culture tubes, shake flasks (using a standard air incubator, or modified bioreactor shake flask attachment), or bioreactor. Growth of parental cells may include supplemented additions to assist regulation of expression constructs listed in the sections above. These supplements may include, but are not limited to dextrose, phosphate, inorganic salts, ribonucleic acids, deoxyribonucleic acids, buffering agents, thiamine, or other chemical that stimulates growth, stabilizes growth, serves as an osmo-protectant, regulates gene expression, and/or applies selective pressure to mutation, and/or marker selection. These mutations may include an amino acid or nucleotide auxotrophy, while the selectable marker may include transposable elements, plasmids, bacteriophage, and/or auxotrophic or antibiotic resistance marker. Growth conditions may also require temperature adjustments, carbon alternations, and/or oxygen-level modifications to stimulate temperature sensitive mutations found in designed gene products for a given desired phenotype and optimize culture conditions.

[0372] By way of non-limiting example, production of minicells and protein production may occur by using either of two general approaches or any combination of each. First, minicells may be formed, purified, and then contained expression elements may be stimulated to produce their encoded gene products. Second, parental cells, from which the minicells are derived, may be stimulated to express the protein of interest and segregate minicells simultaneously. Finally, any timing variable of minicell formation and protein production may be used to optimize protein and minicell production to best serve the desired application. The two general approaches are shown in the sections below.

[0373] III.C. Manipulation of Genetic Expression in Minicell Production

[0374] Included in the design of the invention are methods that increase the efficiency, rate and/or level of gene expression and protein production in parent cells and/or minicells. Such methods include, but are not limited to, the following.

[0375] By way of non-limiting example, parental cells are grown overnight in the appropriate media. From this culture, the cells are subcultured into the same media and monitored for growth. At the appropriate cell density, the cultures are induced for minicell production using any of the switching mechanisms discussed in section II.B, regulating any construct discussed in section II.A. Non-limiting examples of this minicell-inducing switching mechanism may be the ileR gene product regulating the production of the ileC minicell-inducing gene product or the mcrR gene product regulating the production of the minB minicell-inducing gene product. Following minicell induction, the culture is allowed to continue growth until the desired concentration of minicells is obtained. At this point, the minicells are separated from the parental cells as described in section II.E. Purified minicells are induced for protein production by triggering the genetic switching mechanism that segregated into the minicell upon separation from the parental cell. By way of non-limiting example, this genetic switching mechanism may be any of those discussed in section I.B, regulating the production of any protein of interest. Furthermore, at this point or during the production of minicells the peripheral gene expression, production, and assembly machinery discussed in section II.C. may be triggered to assist in this process. By way of non-limiting example, the groEL complex may be triggered using the temperature sensitive lambda cI inducible system from a co-segregant plasmid to assist in the proper assembly of the expressed protein of interest.

[0376] III.D. Separation of Minicells from Parent Cells

[0377] A variety of methods are used to separate minicells from parent cells (i.e., the cells from which the minicells are produced) in a mixture of parent cells and minicells. In general, such methods are physical, biochemical and genetic, and can be used in combination.

[0378] III.D.1. Physical Separation of Minicells from Parent Cells


Some techniques involve different centrifugation techniques, e.g., differential and zonal centrifugation. By way of non-limiting example, minicells may be purified by the double sucrose gradient purification technique described by Frazer and Curtiss, Curr. Topics Microbiol. Immunol. 69:1-84, 1975. The first centrifugation involves differential centrifugation, which separates parent cells from minicells based on differences in size and/or density. The percent of sucrose in the gradient (graduated from about 5 to about 20%), Ficoll or glycerol is designed to allow only parent cells to pass through the gradient.

The supernatant, which is enriched for minicells, is then separated from the pellet and is spun at a much higher rate (e.g., 11,000 g). This pelleted the minicells and any parent cells that did not pellet out in the first spin. The pellet is then resuspended and layered on a sucrose gradient.

The band containing minicells is collected, pelleted by centrifugation, and loaded on another gradient. This procedure is repeated until the minicell preparation is essentially depleted of parent cells, or has a concentration of parent cells that is low enough so as not to interfere with a chosen minicell application or activity. By way of non-limiting example, buffers and media used in these gradient and resuspension steps may be LB, defined minimal media, e.g. M63 salts with defined carbon, nitrogen, phosphate, magnesium, and sulfate levels, complex minimal media, e.g. defined minimal media with casamino acid supplement, and/or other buffer or media that serves as an osmo-protectant, stabilizing agent, and/or energy source, or may contain agents that limit the growth of contaminating parental cells, e.g azide, antibiotic, or lack an auxotrophic supplemental requirement, e.g. thiamine.

Other physical methods may also be used to remove parent cells from minicell preparations. By way of non-limiting example, mixtures of parent cells and minicells are frozen to –20°C. and then thawed slowly (Frazer and Curtiss, Curr. Topics Microbiol. Immunol. 69:1-84, 1975).

III.D.2. Biochemical Separation of Minicells from Parent Cells

Contaminating parental cells may be eliminated from minicell preparations by incubation in the presence of an agent, or under a set of conditions, that selectively kills dividing cells. Because minicells can neither grow nor divide, they are resistant to such treatments.

Examples of biochemical conditions that prevent or kill dividing parental cells is treatment with a antibacterial agent, such as penicillin or derivatives of penicillin. Penicillin has two potential affects. First, penicillin prevent cell wall formation and leads to lysis of dividing cells. Second, prior to lysis dividing cells form filaments that may assist in the physical separation steps described in section III.E.1. In addition to penicillin and its derivatives, other agents may be used to prevent division of parental cells. Such agents may include azide. Azide is a reversible inhibitor of electron transport, and thus prevents cell division. As another example, D-cycloserine or phage MS2 lysis protein may also serve as a biochemical approach to eliminate or inhibit dividing parental cells. (Markiewicz et al., FEBS Microbiol. Lett. 70:119-123, 1992). Khachatourians (U.S. Pat. No. 4,311,797) states that it may be desirable to incubate minicell/parent cell mixtures in brain heart infusion broth at 36°C. to 38°C. prior to the addition of penicillin G and further incubations.

III.D.3. Genetic Separation of Minicells from Parent Cells

Alternatively or additionally, various techniques may be used to selectively kill, preferably lyse, parent cells. For example, although minicells can internally retain M13 phage in the plasmid stage of the M13 life cycle, they are refractory to infection and lysis by M13 phage (Staudenbauer et al., Mol. Gen. Genet. 138:203-212, 1975). In contrast, parent cells are infected and lysed by M13 and are thus not selectively removed from a mixture comprising parent cells and minicells. A mixture comprising parent cells and minicells is treated with M13 phage at an M.O.I.—5 (phage-cells). The infection is allowed to continue to a point where ≥50% of the parent cells are lysed, preferably ≥75%, more preferably ≥95% most preferably ≥99%; and ≥25% of the minicells are lysed or killed, preferably ≥15%, most preferably ≥1%.

As another non-limiting example of a method by which parent cells can be selectively killed, and preferably lysed, a chromosome of a parent cell may include a conditionally lethal gene. The induction of the chromosomal lethal gene will result in the destruction of parent cells, but will not affect minicells as they lack the chromosome harboring the conditionally lethal gene. As one example, a parent cell may contain a chromosomal integrated bacteriophage comprising a conditionally lethal gene. One example of such a bacteriophage is an integrated lambda phage that has a temperature sensitive repressor gene (e.g., lambda cI857). Induction of this phage, which results in the destruction of the parent cells but not the a chromosomal minicells, is achieved by simply raising the temperature of the growth media. A preferred bacteriophage to be used in this method is one that kills and/or lyses the parent cell but does not produce infective particles. One non-limiting example of this type of phage is one that lyses a cell but which has been engineered to not produce capsid proteins that are surrounding and protect phage DNA in infective particles. That is, capsid proteins are required for the production of infective particles.

As another non-limiting example of a method by which parent cells can be selectively killed or lysed, toxic proteins may be expressed that lead to parental cell lysis. By way of non-limiting example, these inducible constructs may employ a system described in section III.B. to control the expression of a phage holing gene. Holin genes fall with in at least 35 different families with no detectable orthologous relationships (Grundling, A., et al. 2001. Holins kill without warning. Proc. Natl. Acad. Sci. 98:9348-9352) of which each and any may be used to lyse parental cells to improve the purity of minicell preparations.

Gram negative eubacterial cells and minicells are bounded by an inner membrane, which is surrounded by a cell wall, wherein the cell wall is itself enclosed within an outer membrane. That is, proceeding from the external environment to the cytoplasm of a minicell, a molecule first encounters the outer membrane (OM), then the cell wall and finally, the inner membrane (IM). In different aspects of the invention, it is preferred to disrupt or degrade the OM, cell wall or IM of a eubacterial minicell. Such treatments are used, by way of non-limiting example, in order to increase or decrease the immunogenicity, and/or to alter the permeability characteristics, of a minicell.
Eubacterial cells and minicells with altered membranes and/or cell walls are called “Poroplasts™” “spheroplasts,” and “protoplasts.” Herein, the terms “spheroplast” and “protoplast” refer to spheroplasts and protoplasts prepared from minicells. In contrast, “cellular spheroplasts” and “cellular protoplasts” refer to spheroplasts and protoplasts prepared from cells. Also, as used herein, the term “minicell” encompasses not only minicells per se but also encompasses Poroplasts™, spheroplasts and protoplasts.

In a poroplast, the eubacterial outer membrane (OM) and LPS have been removed. In a spheroplast, portions of a disrupted eubacterial OM and/or disrupted cell wall either may remain associated with the inner membrane of the minicell, but the membrane and cell wall is nonetheless porous because the permeability of the disrupted OM and cell wall has been increased. A membrane is said to be “disrupted” when the membrane’s structure has been treated with an agent, or incubated under conditions, that leads to the partial degradation of the membrane, thereby increasing the permeability thereof. In contrast, a membrane that has been “degraded” is essentially, for the applicable intents and purposes, removed. In preferred embodiments, irrespective of the condition of the OM and cell wall, the eubacterial inner membrane is not disrupted, and membrane proteins displayed on the inner membrane are accessible to compounds that are brought into contact with the minicell, poroplast, spheroplast, protoplast or cellular poroplast, as the case may be.

For various applications poroplastic minicells are capable of preserving the cytoplasmic integrity while producing increased stability over that of naked protoplasts. Maintenance of the cell wall in poroplastic minicells increases the osmotic resistance, mechanical resistance and storage capacity over protoplasts while permitting passage of small and medium size proteins and molecules through the porous cell wall. A poroplast is a Gram negative bacterium that has its outer membrane only removed. The production of poroplastic involves a modification of the procedure to make protoplasts to remove the outer membrane (Birdsell et al., Production and ultrastructure of lysozyme and ethylenediaminetetraacetic acid-lysozyme Spheroplasts of Escherichia coli, J. Bacteriology 93: 427-437, 1967; Weiss, Protoplast formation in Escherichia coli, J. Bacteriology. 128:668-670, 1976). Like protoplasts, measuring the total LPS that remains in the poroplast preparation may be used to monitor the removal of the outer membrane. Endotoxin kits and antibodies reactive against LPS may be used to measure LPS in solution; increasing amounts of soluble LPS indicates decreased retention of LPS by protoplasts. This assay thus makes it possible to quantify the percent removal of total outer membrane from the poroplastic minicells.


III.E.3. Spheroplasts

A spheroplast is a bacterial minicell that has a disrupted cell wall and/or a disrupted OM. Unlike eubacterial minicells and poroplasts, which have a cell wall and can thus retain their shape despite changes in osmotic conditions, the absence of an intact cell wall in spheroplasts means that these minicells do not have a rigid form.

III.E.4. Protoplasts

A protoplast is a bacterium that has its outer membrane and cell wall removed. The production of protoplasts involves the use of lysozyme and high salt buffers to remove the outer membrane and cell wall (Birdsell et al., Production and ultrastructure of lysozyme and ethylenediaminetetraacetate-lysozyme Spheroplasts of Escherichia coli, J. Bacteriology 93: 427-437, 1967; Weiss, Protoplast formation in Escherichia coli, J. Bacteriology. 128:668-670, 1976). Various commercially available lysozymes can be used in such protocols. Measuring the total LPS that remains in the protoplast preparation is used to monitor the removal of the outer membrane. Endotoxin kits assays can be used to measure LPS in solution; increasing amounts of soluble LPS indicates decreased retention of LPS by protoplasts. This assay thus makes it possible to quantify the percent removal of total outer membrane from the minicells. Endotoxin assays are commercially available from, e.g., BioWhittaker Molecular Applications (Rockland, Me.)

For minicell applications that utilize bacterial-derived minicells, it may be necessary to remove the outer membrane of Gram-negative cells and/or the cell wall of any bacterial-derived minicell. For Gram-positive bacterial cells, removal of the cell wall may be easily accomplished using lysozyme. This enzyme degrades the cell wall allowing easy removal of soluble cell wall components from the pelletable protoplasted minicells. In a more complex system, the cell wall and outer membrane of Gram-negative cells may be removed by combination treatment with EDTA and lysozyme using a step-wise approach in the presence of an osmoprotecting agent (Birdsell, et al. 1967. Production and ultrastructure of lysozyme and ethylenediaminetetraacetate-lysozyme spheroplasts of E. coli, J. Bacteriol. 93:427-437; Weiss, 1976, Protoplast formation in E. coli, J. Bacteriol. 128:668-670). By non-limiting By way of non-limiting example, this osmoprotectant may be sucrose and/or glycerol. It has been found that the concentration of the osmoprotectant sucrose, the cell wall digesting enzyme lysozyme, and chelator EDTA can be opti-
imized to increase the quality of the protoplasts produced. Separation of either prepared Gram-negative spheroplasts prepared in either fashion from removed remaining LPS may occur through exposure of the spheroplast mixture to an anti-LPS antibody. By non-limiting By way of non-limiting example, the anti-LPS antibody may be covalently or non-covalently attached to magnetic, agarose, sepharose, sephaeryl, polyacrylamide, and/or sephadex beads. Following incubation, LPS is removed from the mixture using a magnet or slow centrifugation resulting in a protoplast-enriched supernatant.

[0402] Monitoring loss of LPS may occur through dot-blot analysis of protoplast mixtures or various commercially available endotoxin kit assays can be used to measure LPS in solution; increasing amounts of soluble LPS indicates decreased retention of LPS by protoplasts. This immunological assay may comprise a step of comparing to a standard curve in order to quantify the percent removal of total outer membrane from the minicells. Other endotoxin assays, such as the LAB Systems from BioWhittaker, are commercially available. LPS removal has been measured by gas chromatography of fatty acid methyl esters. Alakomi H.L., Skyttä E., Saarela M., Mattila-Sandholm T., Latva-Kala K.,Helander M., Lactic acid permeabilizes gram-negative bacteria by disrupting the outer membrane. Appl Environ Microbiol. 2000 May; 66(5):2001-5.

[0403] In order to reduce, preferably eliminate, in vivo antigenic potential of minicells or minicell protoplasts, minicell protoplasts may be treated to remove undesirable surface components. Minicell protoplasts so treated are referred to as “denuded minicells” a term that encompasses both spheroplasts and protoplasts. Denuding minicells or minicell protoplasts is accomplished by treatment with one or more enzymes or conditions that selectively or preferentially removes or makes less antigenic externally displayed proteins. As one non-limiting example, the protease trypsin is used to digest exposed proteins on the surface of these particles. In this example, the proteolytic activity of trypsin may be modified or terminated by the addition of a soybean trypsin inhibitor. Non-limiting examples of other proteases that additionally or alternatively may be used include chymotrypsin, papain, elastase, proteinase K and pepsin. For some such proteases, it may be necessary to limit the extent of proteolysis by, e.g., using a suboptimal concentration of protease or by allowing the reaction to proceed for a suboptimal period of time. By the term “suboptimal,” it is meant that complete digestion is not achieved under such conditions, even though the reactions could proceed to completion under other (i.e., optimal) conditions.

[0404] It is sometimes preferred to use molecular genetic techniques to create mutant derivatives of exogenous proteins that (1) are resistant to the proteases or other enzymes used to prepare minicells and (2) retain the desired biological activity of the receptor that is desired to be retained, i.e., the ability to bind one or more ligands of interest.

[0405] It is within the scope of the invention to have two or more exogenous proteins expressed within and preferentially displayed by minicells in order to achieve combined, preferable synergistic, therapeutic compositions. Similarly, two or more therapeutic minicell compositions are formulated into the same composition, or are administered simultaneously with the same therapeutic minicell compositions (i.e., “cocktail” therapies). In other types of “cocktail” therapy, one or more therapeutic minicell compositions are combined or co-administered with one or more other therapeutic agents that are non-minicell compositions such as, e.g., organic compounds, therapeutic proteins, gene therapy constructs, and the like.

[0406] III.F. Minicells from L-Form Eubacteria

[0407] L-form bacterial strains may be used to prepare minicells and are preferred in some embodiments of the invention. L-form bacterial strains are mutant or variant strains, or eubacteria that have been subject to certain conditions, that lack an outer membrane, a cell wall, a periplasmic space and extracellular proteases. Thus, in L-form Eubacteria, the cytoplasmic membrane is the only barrier between the cytoplasm and its surrounding environment. For reviews, see Grischko, V. P., et al. 1999. The Potential of L-Form Bacteria in Biotechnology. Can. J. Chem. Engineering 77:973-977; and Gunpertt, J., et al. 1998 Use of cell wall-less bacteria (L-forms) for efficient expression and secretion of heterologous gene products. Curr Opin Biotechnol. 9:506-9.

[0408] Segregation of minicells from L-form eubacterial parent cells allows for the generation of minicells that are at least partially deficient in barriers that lie outside the cytoplasmic membrane, thus providing direct access to components displayed on the minicell membrane. Thus, depending on the strains and methods of preparation used, minicells prepared from L-form eubacterial parent cells will be similar if not identical to various forms of poroplasts, spheroplasts and/or protoplasts. Displayed components that are accessible in L-form minicells include, but are not limited to, lipids, small molecules, proteins, sugars, nucleic acids and/or moieties that are covalently or non-covalently associated with the cytoplasmic membrane or any component thereof.

IV.B. Detecting Protein Synthesis in Minicells

The level of minicell production will vary and may be evaluated using methods described herein. Relatively high levels of minicell production are generally preferred. Conditions in which about 40% of cells are achronosomal have been reported (see, e.g., Hassan et al., Suppression of initiation defects of chromosome replication in Bacillus subtilis dnaA and oriC-deleted mutants by integration of a plasmid replicon into the chromosones, J. Bacteriol. 179:2494-502, 1997). Procedures for identifying strains that give high yields of minicells are known in the art; see, e.g., Clark-Curtiss and Curtiss III, Analysis of Recombinant DNA Using Escherichia coli Minicells, Meth. Enzol. 101:347-362, 1983.

Minicell production can be assessed by microscopic examination of late log-phase cultures. The ratio of minicells to normal cells and the frequency of cells actively producing minicells are parameters that increase with increasing minicell production.

IV.C. Molecular Sponge

Minicell compositions can also be tested for their ability to bind and/or internalize toxic compounds. The therapeutic potential of such capacity is evaluated using experiments in which detectably labeled derivatives of a toxic compound are present in the bloodstream of an anesthetized animal, which may a human. The blood of the animal is shunted out of the body and past a device that incorporates a minicell composition before being returned to the body. The device is constructed so that the blood contacts a semipermeable membrane that is in contact with the minicell composition. By “semipermeable” it is meant that certain agents can be freely exchanged across the membrane, whereas others are retained on one side of the membrane or the other. For example, the toxic compound of interest is able to cross the semipermeable membrane, whereas minicells and blood cells are separately retained in their respective compartments. Detectably labeled derivatives of the toxic compound are present in the bloodstream of the animal. The capacity of the minicells to take up the toxic compound corresponds with a reduction of the levels of detectably labeled material in the blood and an increase in detectably labeled material in the minicell composition.
[0424] The above types of minicell-comprising compositions, devices, and procedures may be incorporated into ex vivo modalities such as ex vivo gene therapy and dialysis machines. An "ex vivo modality" is one in which a biological sample, such as a blood sample, is temporarily removed from an animal, altered through in vitro manipulation, and then returned to the body. In "ex vivo gene therapy," cells in the sample from the animal are transformed with DNA in vitro and then returned to the body. A "dialysis machine" is a device in which a fluid such as blood of an animal is temporarily removed therefrom and processed through one or more physical, chemical, biochemical, binding or other processes designed to remove undesirable substances including but are not limited to toxins, venoms, overexpressed or overactive endogenous agents, and pathogens or molecules derived therefrom.

[0425] Intraminicellular co-expression of a second molecule that is displayed on the surface of minicells, and which is a ligand for a binding moiety that is immobilized, can optionally be used in order to remove minicells from the sample before it is returned to the body. In the latter aspect, minicells that bind undesirable substances are preferably removed with the undesirable compound remaining bound to the minicells. Minicells that have been used for ex vivo gene therapy, but which have failed to deliver a nucleic acid to any cells in the sample, can be removed in a similar manner.

[0426] IV.C.3. Minicell-Solubilized Receptors

[0427] It is known in the art to use recombinant DNA technology to prepare soluble (hydrophilic) receptor fragments from receptors that bind a bioactive ligand. Unlike the native, membrane-bound receptor, which is relatively insoluble in water (hydrophobic), soluble receptor fragments can be formulated for therapeutic delivery using techniques that are known to have been used to formulate soluble agents.

[0428] Typically, soluble receptor fragments are used to competitively inhibit the binding of the receptor to its ligand. That is, the soluble receptor fragments bind the ligand at the expense of the membrane-bound receptor. Because less of the ligand is bound to its receptor, the cellular response to the ligand is attenuated. Common cellular responses that are desirably attenuated include but are not limited to the uptake of an undesirable agent (e.g., a toxin, a pathogen, etc.) and activation of a signaling pathway having undesirable consequences (e.g., inflammation, apoptosis, unregulated growth, etc.).

[0429] Preparing a soluble fragment derived from a receptor is not trivial. Typically, the three dimensional structure of the receptor is not known, and must be predicted based on homology with other receptors or by using software that predicts the tertiary structure of a polypeptide based on its amino acid sequence. Using the hypothetical structure of the receptor, a series of polypeptides are prepared that comprise amino acid sequences from the receptor but lack regions thereof that are thought to be membrane-anchoring or transmembrane domain(s) of the receptor. Some of the polypeptides prepared this way may be soluble, some may retain the binding activity of the receptor, and a few may have both characteristics. Members of the latter class of polypeptides are soluble receptor fragments, some of which may be amenable to development as a therapeutic or diagnostic agent.

[0430] For any given receptor, there is always the possibility that none of the soluble fragments derived from the receptor will specifically bind to its ligand with sufficient affinity as to be therapeutically effective. Thus, in some instances, it may not be possible to prepare a receptor fragment that is both soluble and sufficiently biologically active.

[0431] The minicells of the invention provide a "universal carrier" for receptors that allows the hydrophilic receptors to be solubilized in the sense that, although they remain associated with a membrane, the minicell is a small, solubilible particle. That is, as an alternative to preparing a set of polypeptides to see which, if any of them, are water soluble receptor fragments, one may, using the teachings of the disclosure, prepare soluble minicells that display the receptor.

[0432] IV.C.4. Reducing Toxicity

[0433] For in vivo use of minicells for the purposes of eliciting an immune response or for therapeutic and diagnostic applications involving delivery of minicells to a human or to an animal, it may be useful to minimize minicell toxicity by using endotoxin-deficient mutants of parent cells. Without being limited to the following example, lipopolysaccharide (LPS) deficient E. coli strains could be conjugated with minicell producing cells to make parent cells lacking the endotoxin. LPS synthesis in E. coli includes the lipid A biosynthetic pathway. Four of the genes in this pathway have now been identified and sequenced, and three of them are located in a complex operon which also contains genes involved in DNA and phospholipid synthesis. The rfa gene cluster, which contains many of the genes for LPS core synthesis, includes at least 17 genes. The rfb gene cluster encodes protein involved in O-antigen synthesis, and rfb genes have been sequenced from a number of serotypes and exhibit the genetic polymorphism anticipated on the basis of the chemical complexity of the O antigens (Schnaitman and Klena, Genetics of lipopolysaccharide biosynthesis in enteric bacteria, Microbiol. Rev. 57:655-82). When present alone or in combination the rfb and rfa mutations cause alterations in the outer membrane that make it more sensitive to lysozyme and other agents used to process minicells. Similarly, the rfa (Chen, L., and W. G. Coleman Jr. 1993. Cloning and characterization of the Escherichia coli K-12 rfa-2 (rfiC) gene, a gene required for lipopolysaccharide inner core synthesis. J. Bacteriol. 175:2534-2540), IpcA (Brooke, J. S., and M. A. Valvano. 1996. Biosynthesis of inner core lipopolysaccharide in enteric bacteria identification and characterization of a conserved phosphoester isomerase. J. Biol. Chem. 271:3608-3614), and IpcB (Kadman, J. L., et al. 1998. Cloning and overexpression of glycosyltransferases that generate the lipopolysaccharide core of Rhizobium leguminosarum. J. Biol. Chem. 273:26432-26440) mutations, when present alone or in combination, cause alterations in lipopolysaccharides in the outer membrane causing cells to be more sensitive to lysozyme and agents used to process minicells. In addition, such mutations can be used to reduce the potential antigenicity and/or toxicity of minicells.

[0434] Minicell-producing cells may comprise mutations that augment preparative steps. For example, lipopolysaccharide (LPS) synthesis in E. coli includes the lipid A biosynthetic pathway. Four of the genes in this pathway have now been identified and sequenced, and three of them are located in a complex operon that also contains genes involved in DNA and phospholipid synthesis. The rfa gene cluster, which contains many of the genes for LPS core synthesis, includes at least 17 genes. The rfb gene cluster encodes protein involved in O-antigen synthesis, and rfb genes have been sequenced from a number of serotypes and exhibit the genetic polymorphism anticipated on the basis of the chemical complexity of the O antigens. See Schnaitman and Klena, Genetics of

**[0435]** V. Genetic Expression in Minicells

**[0436]** Various minicells of the invention use recombinant DNA expression systems to produce a non-endobacterial protein, which may be a membrane protein that is preferably “displayed” on the surface of minicells, a membrane protein that projects portions not associated with a membrane towards the exterior of a minicell, or a soluble protein present in the exterior of the minicells. By “displayed” it is meant that a protein is present on the surface of a cell (or minicell) and is thus in contact with the external environment of the cell. Non-limiting examples of displayed exogenous proteins of the invention include mammalian receptors and fusion proteins comprising one or more transmembrane domains. In other aspects of the invention, minicells use expression elements to produce bioactive nucleic acids from templates theretofor.

**[0437]** V.A. Expression Systems


**[0440]** Preferred expression vectors and constructs according to the invention are episomal genetic elements. By “episomal” it is meant that the expression construct is not always linked to a cell’s chromosome but may instead be retained or maintained in host cells as a distinct molecular entity. Minicells can retain, maintain and express episomal expression constructs such as, e.g., plasmids, bacteriophage, viruses and the like (Crooks et al., Plasmid 10:66-72, 1983; Clark-Curtiss, Methods Enzymology 101:347-362, 1983; Witkiewicz et al., Acta Microbiol. Pol. A 7:21-24, 1975; Ponta et al., Nature 269:440-2, 1977). By “retained” it is meant that the episomal expression construct is capable of autonomous replication within a host parent cell and/or minicell. In the context of episomal elements, the term “contained” encompasses both “retained” and “maintained.” A preferred type of episomal element according to the invention is one that is always an extrachromosomcal element, or which is part of a chromosome but becomes an extrachromosomal element before or during minicell production.

**[0441]** The fact that minicells do not contain chromosomal DNA but do contain episomal expression elements, such as plasmids, that can be used as templates for RNA synthesis means that the only proteins that are actively produced in minicells are those that are encoded by the expression elements that they contain. Minicell-producing *E. coli* cells can be made competent and transformed with expression elements that direct the expression of proteins encoded by the expression elements. An expression element segregates into minicells as they are produced. In isolated minicells that contain expression elements, there is a single DNA template RNA for transcription. Therefore, the only nucleic acids and proteins that are actively produced (expressed) by minicells are those that are encoded by sequences on the expression vector. In the context of the invention, sequences that encode amino acid sequences are designated “open reading frames” or “ORFs.” One feature of minicell expression systems of interest as regards the present invention is that endogenous (i.e., chromosomally located) genes are not present and are thus not expressed, whereas genes present on the episomal element are expressed (preferably over-expressed) in the minicells. As a result, the amount of endogenous proteins, including membrane proteins, decreases as the minicells continue to express genes located on episomal expression constructs.

**[0442]** The minicell system can reduce or eliminate undesirable features associated with the transcription and translation of endogenous proteins from the *E. coli* chromosome. For example, expression of proteins in minicell systems results in low background signal (“noise”) when radiolabeled proteins produced using recombinant DNA technology (Janatpour et al., Translocation of *Vibrio Harveyi* N,N-Diaceetylchitobiose to the outer membrane of *Escherichia coli*. J. Bacteriol, 1987: 169: 3785-3791). A high background signal can make it difficult to detect a protein of interest. In whole cell *E. coli* systems, endogenous proteins (encoded by the bacterial chromosome) are labeled as well as the protein(s) encoded by the expression element; whereas, in minicell systems, only the proteins encoded by the expression element in the minicells are labeled.

**[0443]** There are a variety of proteins, both extrachromosomal and chromosomal, that have been expressed from plasmid DNA in minicells (Clark-Curtiss, Methods Enzymol, 101:347-362, 1983). Some examples of proteins and nucleic acids that have been expressed in minicells include the Kdp-ATPase of *E. coli* (Alldendorf et al., Structure and function of the Kdp-ATPase of *Escherichia coli*. Acta Physiol Scand, 643: 137-

**VI. Fusion (Chimeric) Proteins**

**[0446]**

In certain aspects of the invention, a fusion protein is expressed and displayed by minicells. One class of fusion proteins of particular interest are those that are displayed on the surface of minicells, e.g., fusion proteins comprising one or more transmembrane domains. Types of displayed fusion proteins of particular interest are, by way of non-limiting example, those that have an extracellular domain that is a binding moiety or by way of non-limiting example, the fusion protein ToxR-PhoA has been expressed in and displayed on the surface of minicells. The ToxR-PhoA fusion protein comprises a polypeptide corresponding to the normally soluble enzyme, alkaline phosphatase, anchored to the minicell membrane by the single transmembrane domain of ToxR (see the Examples). The fusion protein retains the activity of the enzyme in the context of the minicell membrane in which it is bound. Nearly all of the fusion protein is oriented so that the enzyme’s catalytic domain is displayed on the outer surface of the minicell.

**[0447]**

**VI.A. Generation of Fusion Proteins**

**[0449]**

Polypeptides, which are polymers of amino acids, are encoded by another class of molecules, known as nucleic acids, which are polymers of structural units known as nucleotides. In particular, proteins are encoded by nucleic acids known as DNA and RNA (deoxyribonucleic acid and ribonucleic acid, respectively).  

**[0450]**

The nucleotide sequence of a nucleic acid contains the “blueprints” for a protein. Nucleic acids are polymers of nucleotides, four types of which are present in a given nucleic acid. The nucleotides in DNA are adenine, cytosine and guanine and thymine, (referred by A, C, G, and T respectively); in RNA, thymine (T) is replaced by uracil (U). The structures of nucleic acids are represented by the sequence of its nucleotides arranged in a 3' ("3 prime") to 5' ("5 prime") direction, e.g.,

\[ 5'\text{-A-T-G-C-C-T-A-A-G-C-C-G-C-T-C-C-T-C-A-3'} \]

**[0451]**

In biological systems, proteins are typically produced in the following manner. A DNA molecule that has a nucleotide sequence that encodes the amino acid sequence of a protein is used as a template to guide the production of a messenger RNA (mRNA) that also encodes the protein; this process is known as transcription. In a subsequent process called translation, the mRNA is “read” and directs the synthesis of a protein having a particular amino acid sequence.  

**[0452]**

Each amino acid in a protein is encoded by a series of three contiguous nucleotides, each of which is known as a codon. In the "genetic code," some amino acids are encoded by several codons, each codon having a different sequence; whereas other amino acids are encoded by only one codon sequence. An entire protein (i.e., a complete amino acid sequence) is encoded by a nucleic acid sequence called a reading frame. A reading frame is a continuous nucleotide sequence that encodes the amino acid sequence of a protein; the boundaries of a reading frame are defined by its initiation (start) and termination (stop) codons.
The process by which a protein is produced from a nucleic acid can be diagrammed as follows:

DNA (A-T-G) - (A-A-G) - (C-C-G) - (C-T-C) - (C-C-T) - ... (etc.) (SEQ ID NO. 362)

Transcription

RNA (A-U-G) - (A-A-G) - (C-C-G) - (C-U-C) - (C-C-U) - ... (etc.) (SEQ ID NO. 363)

Translation

Protein Met - Pro - Lys - Ala - Ala - ... (etc.) (SEQ ID NO. 364)

A chimeric reading frame encoding a fusion protein is prepared as follows. A “chimeric reading frame” is a genetically engineered reading frame that results from the fusion of two or more normally distinct reading frames, or fragments thereof, each of which normally encodes a separate polypeptide. Using recombinant DNA techniques, a first reading frame that encodes a first amino acid sequence is linked to a second reading frame that encodes a second amino acid sequence in order to generate a chimeric reading frame. Chimeric reading frames may also include nucleotide sequences that encode optional fusion protein elements (see below).

A hypothetical example of a chimeric reading frame created from two normally separate reading frames is depicted in the following flowchart.

First Open Reading Frame and “Protein-1”:

DNA-1 (A-T-G) - (A-A-G) - (C-C-G) - (C-T-C) - (C-C-T) - ... (etc.) (SEQ ID NO. 362)

Transcription

RNA-1 (A-U-G) - (A-A-G) - (C-C-G) - (C-U-C) - (C-C-U) - ... (etc.) (SEQ ID NO. 363)

Translation

Protein-1 Met-Pro-Lys-Ala-Ala- ... (etc.) (SEQ ID NO. 364)

Second Open Reading Frame and “Protein-2”:

DNA-2 (T-G-G) - (G-T-T) - (A-C-T) - (C-A-C) - (T-C-A) - ... (etc.) (SEQ ID NO. 365)

Transcription

RNA-2 (U-G-G) - (G-U-U) - (A-C-U) - (C-A-C) - (U-C-A) - ... (etc.) (SEQ ID NO. 366)

Translation

Protein-2 Trp-Val-Thr-His-Ser- ... (etc.) (SEQ ID NO. 367)
Chimeric Reading Frame that Encodes a Fusion Protein Having Sequences from Protein-1 and Protein-2:

DNA-Chimera (A-T-G) - (A-A-G) - (C-C-G) - (C-A-C) - (T-C-A) - (etc.) (SEQ ID NO. 368)

Transcription

DNA-Chimera (A-U-G) - (A-A-G) - (C-C-G) - (C-A-C) - (U-C-A) - (etc.) (SEQ ID NO. 369)

Transcription

Fusion Protein Met-Pro-Lys-His-Ser-(etc.) (SEQ ID NO. 370)

In order for a chimeric reading frame to be functional, each normally distinct reading frame therein must be fused to all of the other normally distinct reading frames in a manner such that all of the reading frames are in frame with each other. By “in frame with each other” it is meant that, in a chimeric reading frame, a first nucleic acid having a first reading frame is covalently linked to a second nucleic acid having a second reading frame in such a manner that the two reading frames are “read” (translated) in register with each other. As a result, the chimeric reading frame encodes one extended amino acid sequence that includes the amino acid sequences encoded by each of the normally separate reading frames. A fusion protein is thus encoded by a chimeric reading frame.

The fusion proteins of the invention are used to display polypeptides on minicells. The fusion proteins comprise (1) at least one polypeptide that is desired to be displayed by minicells (a “displayed polypeptide”) and (2) at least one membrane polypeptide, e.g., a transmembrane or a membrane anchoring domain. For various aspects of the invention, optional fusion protein elements, as defined herein, may also be included if required or desired.

VI.B. Optional Fusion Protein Elements

The fusion proteins of the invention may optionally comprise one or more non-biologically active amino acid sequences, i.e., optional fusion protein elements. Such elements include, but are not limited to, the following optional fusion protein elements. It is understood that a chimeric reading frame will include nucleotide sequences that encode such optional fusion protein elements, and that these nucleotide sequences will be positioned so as to be in frame with the reading frame encoding the fusion protein. Optional fusion protein elements may be inserted between the displayed polypeptide and the membrane polypeptide, upstream or downstream (amino proximal and carboxy proximal, respectively) of these and other elements, or within the displayed polypeptide and the membrane polypeptide. A person skilled in the art will be able to determine which optional element(s) should be included in a fusion protein of the invention, and in what order, based on the desired method of production or intended use of the fusion protein.

Detectable polypeptides are optional fusion protein elements that either generate a detectable signal or are specifically recognized by a detectably labeled agent. An example of the former class of detectable polypeptide is green fluorescent protein (GFP). Examples of the latter class include epitopes such as a “His tag” (6 contiguous His residues, a.k.a. 6xHis), the “FLAG tag” and the e-myc epitope. These and other epitopes can be detected using labeled antibodies that are specific for the epitope. Several such antibodies are commercially available.

Attachment (support-binding) elements are optionally included in fusion proteins and can be used to attach minicells displaying a fusion protein to a preselected surface or support. Examples of such elements include a “His tag,” which binds to surfaces that have been coated with nickel; streptavidin or avidin, which bind to surfaces that have been coated with biotin or “biotinylated” (see U.S. Pat. No. 4,839,293 and Airemen et al., Protein Expr. Purif. 17:139-145, 1999); and glutathione-SS-transferase (GST), which binds to surfaces coated with glutathione (Kaplan et al., Proteins Sci. 6:399-406, 1997; U.S. Pat. No. 5,654,176). Polypeptides that bind to lead ions have also been described (U.S. Pat. No. 6,111,079).

Spacers (a.k.a. linkers) are amino acid sequences that are optionally included in a fusion protein in between other portions of a fusion protein (e.g., between the membrane polypeptide and the displayed polypeptide, or between an optional fusion protein element and the remainder of the fusion protein). Spacers can be included for a variety of reasons. For example, a spacer can provide some physical separation between two parts of a protein that might otherwise interfere with each other via, e.g., steric hindrance. The ability to manipulate the distance between the membrane polypeptide and the displayed polypeptide allows one to extend the displayed polypeptide to various distances from the surface of minicells.

VI.C. Interactions with Recipient Cells

Many Gram-negative pathogens use a type III secretion machine to translocate protein toxins across the bacterial cell envelope (for a review, see Cheng L W, Schneewind O. Type III machines of Gram-negative bacteria: delivering the goods. Trends Microbiol 2000 May; 8(5):214-20). For example, pathogenic Yersinia spp. export over a dozen Yop proteins via a type III mechanism, which recognizes secretion substrates by signals encoded in yop mRNA or chaperones bound to unfolded Yop proteins. A 70-kb virulence plasmid found in pathogenic Yersinia spp. to survive and multiply in the lymphoid tissues of the host. The virulence plasmid encodes the Yop virulin, an integrated system allowing extracellular bacteria to inject bacterial proteins into cells. The Yop virulin comprises a variety of Yop proteins and a dedicated type III secretion apparatus, called Ysc (for a review, see Cornelis G R, Boland A, Boyd A P, Gruenert C, Iriarte M, Neyt...

VII. Minicell Display

Included in the design of the invention is the use of minicells to express and display soluble or membrane-bound protein libraries to identify a soluble or membrane-bound protein that binds a known ligand or to identify proteins (e.g., orphan receptors) for which the known ligand or substrate is not known but for which a reporter could be engineered into the minicell that would signal the presence of the encoded protein. In the preferred embodiment of the invention, this ‘minicell display’ technique is analogous to plaque display for the purpose of identifying genes that encode receptor-like or antibody-like proteins against known ligand. This approach will allow identification of an unknown receptor protein for which a known ligand has affinity. These known ligands may have been identified as having a pharmacological, biological, or other effect without knowledge of the site of effect. In these cases the knowledge of receptor will allow basic research to understand the molecular and/or physiological response and permit directed modification of the ligand for better pharmacological or biological response or modification of the receptor for employment in ligand-binding applications. In another non-limiting embodiment of the invention, the ligand need not be known but some general characteristic of the protein would be.

For purposes of this application, soluble or membrane-bound protein libraries may be constructed by random cloning of DNA fragments or directed cloning using reverse transcriptase polymerase chain reaction (RT-PCR). In either method, DNA fragments may be placed under the regulation of any regulatory element listed in section II.B. on any plasmid or chromosomal construct. In the case of soluble protein receptors, they will be fused to form a chimeric protein with a known transmembrane domain (TMD), e.g. the TMD from the toxR gene product. Upon induction of the soluble or membrane-bound protein library, minicells, minicell protoplasts, or minicell poroplasts (as the experiment requires) will be mixed with the known ligand. Without being limited to the following example, screening could be accomplished by first labeling the known ligand with a molecular fluorophore, e.g. TAMRA, FTR, or in some cases a fluorescent protein, e.g. GFP. A positive interaction between the minicells displaying the receptor for the labeled ligand will be identified and separated from the library population by fluorescent-activated cell sorting (FACS). Isolated, positive receptor-ligand interactions will be identified by PCR amplification, subcloned into a clean background, and sequenced using plasmid-specific oligonucleotides. Subcloned proteins will be re-screened for interaction with the labeled ligand, and their binding patterns characterized.

Positive interacting receptor proteins may be employed in mutagenesis or other directed evolutionary process to improve or decrease the binding affinity to the ligand. In another application, the receptor-ligand pair may be further employed in a screening process to identify new compounds that may interfere with the interaction. Thus, using a known substance to identify the receptor and the identified receptor-ligand pair to identify other interfering compounds. Chimeric-soluble or membrane-bound protein libraries may be screened versus a protein-array chip that presents a variety of known protein compounds or peptide variations. In this application, the minicell, minicell protoplast, or minicell poroplast will also contain a label, signaling component, and/or antigen recognizable by an antibody for identification of a positive interaction on the protein chip array. Other approaches for identification may include packaged fluorescent molecules or proteins that are constitutively produced, induced by the positive interaction with the ligand, or regulated by a regulatory element described in section II.B.

In a preferred embodiment of the invention, cDNA libraries could be constructed from isolated B-cells, activated B-cell or T-cells for the purpose of identifying receptors or antibodies that are encoded by these cells of the immune system. In a non-limiting example, a small molecule could be used to immunize an experimental animal (e.g., rat, mouse, rabbit), the spleen could be removed, or blood could be drawn and used as a source of mRNA. Reverse transcription reactions could then be used to construct a cDNA library that would eventually be transformed into the minicell parent bacteria, as described above. The minicells would then be isolated, induced and subjected to FACS analysis with subsequent amplification and sequencing of the cDNA fragment of interest (see above). The PCR-amplified plasmid-containing cDNA fragment encoding the ‘receptor’ or ‘antibody’ of interest would be ready for transformation and expression in the minicell context for diagnostic, therapeutic research or screening applications of the invention.

In a related, non-limiting embodiment of the invention, minicells expressing a particular antigen (e.g., protein, carbohydrate, small molecule, lipid) on their surfaces (described elsewhere in this application) are used to generate an immunogenic response. The advantages of presenting an antigen on the surfaces of minicells are that the minicells themselves may be an adjuvant that stimulates the immune response, particularly if administered subcutaneously (SC) or intramuscularly (IM). Moreover, the minicells are not readily eliminated by the renal system and are present in the circulatory system of an immunized animal for a longer time. In addition, small molecules could be tethered to the minicell in a way that presents the desired moiety of the molecule. Animals are presented with minicell-based immunogens, and the antibodies produced in the animals are prepared and used in therapeutic, diagnostic, research and screening applications. Although this aspect of the invention may be used to make antibodies to any molecule displayed on their surface, the extracellular domains of membrane proteins are of particular interest.

Minicell display could be used to identify orphan receptors or other proteins for which a ligand or substrate is not known. As a non-limiting example, orphan G protein coupled receptors (GPCRs) or novel RNA and DNA polymerses could be identified from organisms living in extreme environments. A cDNA library could be constructed from an organism and expressed in minicells that co-express a reporter system that indicates the presence of the novel protein. In a non-limiting example of GPCRs, the minicells used for minicell display are engineered to express a G-protein in a manner that would signal an interaction with the orphan GPCR.

VIII. Aptamers

Traditionally, techniques for detecting and purifying target molecules have used polypeptides, such as antibodies, that specifically bind such targets. While nucleic acids have long been known to specifically bind other nucleic acids (e.g., ones having complementary sequences), aptamers (i.e., nucleic acids that bind non-nucleic target molecules) have

[0476] As applied to aptamers, the term “binding” specifically excludes the “Watson-Crick”-type binding interactions (i.e., A:T and G:C base-pairing) traditionally associated with the DNA double helix. The term “aptamer” thus refers to a nucleic acid or a nucleic acid derivative that specifically binds to a target molecule, wherein the target molecule is either (i) not a nucleic acid, or (ii) a nucleic acid or structural element thereof that is bound through mechanisms other than duplex- or triplex-type base pairing. Such a molecule is called a “non-nucleic molecule” herein.

[0477] VIII.A. Structures of Nucleic Acids

[0478] “Nucleic acids,” as used herein, refers to nucleic acids that are isolated a natural source; prepared in vitro, using techniques such as PCR amplification or chemical synthesis; prepared in vivo, e.g., via recombinant technology; or by any appropriate method. Nucleic acids may be of any shape (linear, circular, etc.) or topology (single-stranded, double-stranded, supercoiled, etc.). The term “nucleic acids” also includes without limitation nucleic acid derivatives such as peptide nucleic acids (PNA’s) and polypeptide-nucleic acid conjugates; nucleic acids having at least one chemically modified sugar residue, backbone, internucleotide linkage, base, nucleoside, or nucleotide analog; as well as nucleic acids having chemically modified 5’ or 3’ ends; and nucleic acids having two or more of such modifications. Not all linkages in a nucleic acid need to be identical.

[0479] Nucleic acids that are aptamers are often, but need not be, prepared as oligonucleotides. Oligonucleotides include without limitation RNA, DNA and mixed RNA-DNA molecules having sequences of lengths that have minimum lengths of 2, 4, 6, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides, and maximum lengths of about 100, 75, 50, 40, 25, or 15 or more nucleotides, respectively. In general, a minimum of 6 nucleotides, preferably 10 nucleotides, more preferably 14 to 20 nucleotides, is necessary to effect specific binding.

[0480] In general, the oligonucleotides may be single-stranded (ss) or double-stranded (ds) DNA or RNA, or conjugates (e.g., RNA-DNA molecules having 5’ and 3’ DNA “clamps”) or hybrids (e.g., RNA:DNA paired molecules), or derivatives (chemically modified forms thereof). However, single-stranded DNA is preferred, as DNA is often less labile than RNA. Similarly, chemical modifications that enhance an aptamer’s specificity or stability are preferred.

[0481] VIII.B. Chemical Modifications of Nucleic Acids

[0482] Chemical modifications that may be incorporated into aptamers and other nucleic acids include, with neither limitation nor exclusivity, base modifications, sugar modifications, and backbone modifications.

[0483] Base modifications: The base residues in aptamers may be other than naturally occurring bases (e.g., A, G, C, T, U, 5MC, and the like). Derivatives of purines and pyrimidines are known in the art; an exemplary but not exhaustive list includes aziridinylcytosine, 4-acetylcyanosine, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylcytosine, 1-methylguanine, 1-methylalanine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine (5MC), N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxymethyl-2-thiouracil, beta-D-mannosyldeoxyguanosine, 5-methoxyuracil, 2-methylthio-N-6-isopentenyladenine, uracil-5-oxoacetic acid methylster, pseudouracil, queosine, 2-thiocyctosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid, and 2,6-diamino-purine. In addition to nucleic acids that incorporate one or more of such base derivatives, nucleic acids having nucleotide residues that are devoid of a purine or a pyrimidine base may also be included in aptamers.

[0484] Sugar modifications: The sugar residues in aptamers may be other than conventional ribose and deoxyribose residues. By way of non-limiting example, substitution at the 2’-position of the furanose residue enhances nucleic acid stability. An exemplary, but not exhaustive list, of modified sugar residues includes 2’-substituted sugars such as 2’-O-methyl-, 2’-O-alkyl, 2’-O-allyl, 2’-S-alkyl, 2’-S-allyl, 2’-fluoro-, 2’-halo, or 2’-azido-ribose, carbocyclic sugar analogs, alphanoanumeric sugars, epimeric sugars such as arabinoose, xyloses or xyloloses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside, ethyl riboside or propylriboside.

[0485] Backbone modifications: Chemically modified backbones include, by way of non-limiting example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphorothriesters, aminooxylphosphorothriesters, methyl and other alkyl phosphonates including 3’-alkylene phosphonates and chiral phosphonates, phosphonates, phosphoramides including 3-amino phosphoramidate and aminooxyphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionooxylphosphorothiesters, and boranophosphates having normal 3’-5’ linkages, 2’-5’ linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3’-5’ to 5’-3’ or 2’-5’ to 5’-2’. Chemically modified backbones that do not contain a phosphorus atom have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages, including without limitation furanofuranolino linkages; siloxane backbones; sulfoxide, sulfide and sultone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyl-eneiminio and methylenehydraxazine backbones; sulfonate and sulfonamide backbones; and amide backbones.

[0486] VIII.C. Preparation and Identification of Aptamers

[0487] In general, techniques for identifying aptamers involve incubating a preselected non-nucleic target molecule with mixtures (2 to 50 members), pools (50 to 5,000 members) or libraries (50 or more members) of different nucleic acids that are potential aptamers under conditions that allow complexes of target molecules and aptamers to form. By “different nucleic acids” it is meant that the nucleotide sequence of each potential aptamer may be different from that of any other member, that is, the sequences of the potential aptamers are random with respect to each other. Randomness can be introduced in a variety of manners such as, e.g., mutagenesis, which can be carried out in vivo by exposing cells harboring a nucleic acid with mutagenic agents, in vitro by chemical treatment of a nucleic acid, or in vitro by biochemical replication (e.g., PCR) that is deliberately allowed to proceed under conditions that reduce fidelity of replication
process; randomized chemical synthesis, i.e., by synthesizing a plurality of nucleic acids having a preselected sequence that, with regards to at least one position in the sequence, is random. By “random at a position in a preselected sequence” it is meant that a position in a sequence that is normally synthesized as, e.g., as close to 100% A as possible (e.g., 5'- C-T-T-A-G-T-3') is allowed to be randomly synthesized at that position (C-T-T-N-G-T; wherein N indicates a randomized position where, for example, the synthesizing reaction contains 25% each of A, T, C and G; or x% A, w% T, y% C and z% G, wherein x+y+w+z=100). In later stages of the process, the sequences are increasingly less randomized and consensus sequences may appear; in any event, it is preferred to ultimately obtain an aptamer having a unique nucleotide sequence.

[0488] Aptamers and pools of aptamers are prepared, identified, characterized and/or purified by any appropriate technique, including those utilizing in vitro synthesis, recombinant DNA techniques, PCR amplification, and the like. After their formation, target-aptamer complexes are then separated from the uncomplexed members of the nucleic acid mixture, and the nucleic acids that can be prepared from the complexes are candidate aptamers (at early stages of the technique, the aptamers generally being a population of a multiplicity of nucleotide sequences having varying degrees of specificity for the target). The resulting aptamer (mixture or pool) is then substituted for the starting aptamer (library or pool) in repeated iterations of this series of steps. When a limited number (e.g., a pool or mixture, preferably a mixture with less than 100 members, most preferably less than 10 members, most preferably 1, of nucleic acids having satisfactory specificity is obtained, the aptamer is sequenced and characterized. Pure preparations of a given aptamer are generated by any appropriate technique (e.g., PCR amplification, in vitro chemical synthesis, and the like).

[0489] For example, Tuerk and Gold (Science (1990) 249: 505-510) disclose the use of a procedure termed “systematic evolution of ligands by exponential enrichment” (SELEX). In this method, pools of nucleic acid molecules that are randomized at specific positions are subjected to selection for binding to a nucleic acid-binding protein (see, e.g., PCT International Publication No. WO 91/19813 and U.S. Pat. No. 5,270,163). The oligonucleotides so obtained are sequenced and otherwise characterized. Kinzler, K. W., et al. (Nucleic Acids Res. (1989) 17:3645-3653) used a similar technique to identify synthetic double-stranded DNA molecules that are specifically bound by DNA-binding polypeptides. Ellington, A. D., et al. (Nature (1990) 346: 818-822) disclose the production of a large number of random sequence RNA molecules and the selection and identification of those that bind specifically to specific dyes such as Cibacron blue.

[0490] Another technique for identifying nucleic acids that bind non-nucleic target molecules is the oligonucleotide combinatorial technique disclosed by Ecker, D. J. et al. (Nuc. Acids Res. 21, 1855 (1993)) known as “synthetic unrandomization of randomized fragments” (SURF), which is based on repetitive synthesis and screening of increasingly simplified sets of oligonucleotide analogue libraries, pools and mixtures (Tuerk et al., Science 249:505, 1990). The starting library consists of oligonucleotide analogues of defined length with one position in each pool containing a known analogue and the remaining positions containing equimolar mixtures of all other analogues. With each round of synthesis and selection, the identity of at least one position of the oligomer is determined until the sequences of optimized nucleic acid ligand aptamers are discovered.

[0491] Once a particular candidate aptamer has been identified through a SURF, SELEX or any other technique, its nucleotide sequence can be determined (as is known in the art), and its three-dimensional molecular structure can be examined by nuclear magnetic resonance (NMR). These techniques are explained in relation to the determination of the three-dimensional structure of a nucleic acid ligand that binds thrombin in Padmanabhan et al., J. Biol. Chem. 24, 17651 (1993); Wang et al., Biochemistry 32, 1899 (1993); and Macaya et al., Proc. Nat'l. Acad. Sci. USA 90, 3745 (1993). Selected aptamers may be resynthesized using one or more modified bases, sugars or backbone linkages. Aptamers consist essentially of the minimum sequence of nucleic acid needed to confer binding specificity, but may be extended on the 5' end, the 3' end, or both, or may be otherwise derivatized or conjugated.

[0492] IX. Polypeptidic Binding Moieties

[0493] A variety of binding moieties can be attached to a minicell of the invention for a variety of purposes. In a preferred embodiment, the binding moiety is directed to a ligand that is displayed by a cell into which it is desired to deliver the therapeutic content of a minicell.

[0494] IX.A. Antibodies and Antibody Derivatives

[0495] The term “antibody” is meant to encompass an immunoglobulin molecule obtained by in vitro or in vivo generation of an immunogenic response, and includes polyclonal, monospecific and monoclonal antibodies, as well as antibody derivatives, e.g. single-chain antibody fragments (scFv). An “immunogenic response” is one that results in the production of antibodies directed to one or more proteins after the appropriate cells have been contacted with such proteins, or polypeptide derivatives thereof, in a manner such that one or more portions of the protein function as epitopes. An epitope is a single antigenic determinant in a molecule. In proteins, particularly denatured proteins, an epitope is typically defined and represented by a contiguous amino acid sequence. However, in the case of non-denatured proteins, epitopes also include structures, such as active sites, that are formed by the three-dimensional folding of a protein in a manner such that amino acids from separate portions of the amino acid sequence of the protein are brought into close physical contact with each other.

[0496] Wildtype antibodies have four polypeptide chains, two identical heavy chains and two identical light chains. Both types of polypeptide chains have constant regions, which do not vary or vary minimally among antibodies of the same class (i.e., IgA, IgM, etc.), and variable regions. Variable regions are unique to a particular antibody and comprise an “antigen binding domain” that recognizes a specific epitope. Thus, an antibody’s specificity is determined by the variable regions located in the amino terminal regions of the light and heavy chains.

[0497] As used herein, the term “antibody” encompasses derivatives of antibodies such as antibody fragments that retain the ability to specifically bind to antigens. Such antibody fragments include Fab fragments (i.e., an antibody fragment that contains the antigen-binding domain and comprises a light chain and part of a heavy chain bridged by a disulfide bond); Fab' (an antibody fragment containing a single anti-binding domain comprising an Fab and an additional portion of the heavy chain through the hinge region); F(ab')2 (two
Fab molecules joined by interchain disulfide bonds in the hinge regions of the heavy chains; the Fab molecules may be directed toward the same or different epitopes; a bisppecific Fab (an Fab molecule having two antigen binding domains, each of which may be directed to a different epitope); a single chain Fab molecule comprising a variable region, a.k.a., a scFv (the variable, antigen-binding determinative region of a single light and heavy chain of an antibody linked together by a chain of about 10 to about 25 amino acids).

[0498] The term “antibody” includes antibodies and antibody derivatives that are produced by recombinant DNA techniques and “humanized” antibodies. Humanized antibodies have been modified, by genetic manipulation and/or in vitro treatment to be more human, in terms of amino acid sequence, glycosylation pattern, etc., in order to reduce the antigenicity of the antibody or antibody fragment in an animal to which the antibody is intended to be administered (Gussow et al., Methods Enzym. 203:99-121, 1991).

[0499] A single-chain antibody (scFv) is a non-limiting example of a binding moiety that may be displayed on mini-cells. Single-chain antibodies are produced by recombinant DNA technology and may be incorporated into fusion proteins. The term “single chain” denotes the fact that scFv’s are found in single polypeptide. In contrast, wildtype antibodies have four polypeptide chains, two identical heavy chains and two identical light chains. Both types of polypeptide chains have constant regions, which do not vary or vary minimally among antibodies of the same class (i.e., IgA, IgM, etc.), and variable regions. An antibody’s specificity is determined by the variable regions located in the amino terminal regions of the light and heavy chains. The variable regions of a light chain and associated heavy chain form an “antigen binding domain” that recognizes a specific epitope. In a single chain antibody, the amino acid sequences of the variable light and variable heavy regions of an antibody are present in one contiguous polypeptide. Methods of producing single chain antibodies are known in the art. See, for example, U.S. Pat. Nos. 4,946,778; 5,260,203; 5,455,030; 5,518,889; 5,534,621; 5,869,620; 6,025,165; 6,027,725 and 6,121,424.

[0500] Antibody derivatives and other polypeptides that are binding moieties can be isolated from protein display libraries, in which a library of candidate binding agents is displayed on a phage or other host that comprises a nucleic acid encoding the protein it displays. Thus, an agent that binds to the target compound can be isolated, and nucleic acid prepared therefrom, providing for the rapid isolation of binding moieties and nucleic acids that can be used to produce them. For reviews, see Benhar I. Biotechnological applications of phage and cell display. Biotechnology Adv. 2001 (19):1-33; FitzGerald K. in vitro display technologies—new tools for drug discovery. Drug Discov Today. 2000 5(6):253-258; and Hoogenboom H R, Chames P. Natural and designer binding sites made by phage display technology. Immunol Today. 2000 August; 21(8):371-8.


**[0505]** IX.B. Non-Catalytic Derivatives of Active Sites of Enzymes

**[0506]** Enzymes bind their substrates, at least transiently, in regions known as “active sites.” It is known in the art that non-catalytic derivatives of enzymes, which bind but do not chemically alter their substrates may be prepared. Non-catalytic enzymes, particularly the mutant active sites thereof, are used to bind substrate molecules.

**[0507]** As a non-limiting example, enzymes from which biologically inactive (non-catalytic) sphingolipid-binding derivatives are obtained. Such derivatives of these enzymes bind their substrate sphingolipid. Sphingosine-1-phosphate (SIP) is bound by non-catalytic derivatives of enzymes having SIP as a substrate, e.g., SIP lyase and SIP phosphatase. Sphingosine (SHP) is bound by non-catalytic derivatives of enzymes having SHP as a substrate, e.g., SHP kinase and ceramidase synthase. Ceramide (CER) is bound by non-catalytic derivatives of enzymes having CER as a substrate, such as, by way of non-limiting example, ceramidase, sphingomyelin synthase, ceramide kinase, and glucosylceramide synthase. Sphingomyelin is bound by non-catalytic derivatives of sphingomyelinase, an enzyme having sphingomyelin as a substrate.

**[0508]** IX.C. Nucleic Acid Binding Domains

**[0509]** Nucleic acid binding polypeptide domains may bind nucleic acids in a sequence-dependent or sequence-independent fashion and/or in a manner that is specific for various nucleic acids having different chemical structures (e.g., single- or double-stranded DNA or RNA, RNA:DNA hybrid molecules, etc.). Non-limiting examples of membrane-based transcription factors and DNA-binding protein include Smad proteins (Miyazono et al., TGF-beta signaling by Smad proteins (Review), Adv Immunol 75:115-57, 2000); SREBP1s (sterol regulatory element binding proteins) (Ye et al., Asparagine-proline sequence within membrane-spanning

[0510] IX.D. Attaching Binding Motieties, or Other Comounds, to Minicells

[0511] Binding compounds or moieties can be chemically attached (conjugated) to minicells via membrane proteins that are displayed on the minicells. The compound to be conjugated to minicells (the "attachable compound") may be any chemical composition, i.e., a small molecule, a nucleic acid, a radioisotope, a lipid or a polypeptide. One type of attachable compound that can be covalently attached to minicells is a binding moiety, e.g., an antibody or antibody derivative. Another non-limiting example of attachable compounds is polyethylene glycol ("PEG"), which lowers the uptake in vivo of minicells by the reticuloendothelial system (RES). Another non-limiting example of creating stealth minicells to avoid the RES is to express proteins or other molecules on the surfaces of minicells whose lipid compositions have been modified, such as anionic lipid-rich minicells.

[0512] By way of non-limiting example, it is possible to prepare minicells that express transmembrane proteins with cysteine moieties on extracellular domains. Linkage of the membrane protein may be achieved through surface cysteiny1 groups by, e.g., reduction with cysteine residues on other compounds to form disulfide bridges (S-S). If appropriate cysteine residues are not present on the membrane protein they may be introduced by genetic manipulation. The substitution of cysteine for another amino acid may be achieved by methods well-known to those skilled in the art, for example, by using methods described in Maniatis, Sambrook, and Fritsch (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989). As a non-limiting example, bioactive lysosphingolipids (e.g., sphingosine, sphingosine-1-phosphate, sphingosylphosphorylcholine) are covalently linked to proteins expressed on the surfaces of minicells such that these bioactive lipids are on the surface of the minicells and accessible for therapeutic or diagnostic uses in vivo or in vitro.

[0513] When the attachable moiety and the membrane protein both have a reduced sulfhydryl group, a homobifunctional cross-linker that contains maleimide, pyridyl disulfide, or beta-alpha-haloacetyl groups may be used for cross-linking. Examples of such cross-linking reagents include, but are not limited to, bismaleimidohexane (BMII1) or 1,4-Di-(3-(2-pyridylidithio)propionamidyl)butane (DPDPB). Alternatively, a heterobifunctional cross-linker that contains a combination of maleimide, pyridyl disulfide, or beta-alpha-haloacetyl groups can be used for cross-linking.

[0514] As another non-limiting example, attachable moieties may be chemically conjugated using primary amines. In these instances, a homobifunctional cross-linker that contains succinimide ester, imidoester, acylazide, or isocyanate groups may be used for cross-linking. Examples of such cross-linking reagents include, but are not limited to: Bis[2-(succinimidooxyoctamethylene)ethyl]sulfone (BSOCOES); Bis[2-(sulfosuccinimidooxyoctamethylene)ethyl]sulfone (sulfo-BSOCOES); Disuccinimide suberate (DSS); Bis-(Sulfo succinimidyl) Suberate (BS3); Disuccionimide glutarate (DSG); Diviohio(succinimidylpropionate) (DSP); Dithiobis(sulfo succinimidylpropionate) (DTSSP); Disuccionimideyl tartrate (sulfo-DST); Dithio-bis-maleimidoetanete (DTME); Disuccinimideyl tartrate (DST); Ethylene glycol bis(succinimidyl succinate) (sulfo-EGS); Dimethyl malonimide.2HCl (DMM); Ethylene glycol bis(succinimidyl succinate) (EGS); Dimethyl succinimide.2HCl (DMSC); Dimethyl adipimide.2HCl (DMA); Dimethyl pimeliminate.2HCl (DMP); and Dimethyl suberimide.2HCl (DMS), and Dimethyl 3,3’-dithiobispropionicimide. 2HCl (DTBP). Heterobifunctional cross-linkers that contains a combination of imidoester or succinimide ester groups may also be used for cross-linking.

[0515] As another non-limiting example, attachable moieties may be chemically conjugated using sulfhydryl and primary amine groups. In these instances, heterobifunctional cross-linking reagents are preferable used. Examples of such cross-linking reagents include, but are not limited to: N-succinimido 3-(2-pyridyldithio)propionamide (DPNP); N-succinimido 6-[3’-(2-pyridyldithio)propionamidyl]hexanoate (sulfo-LC-SPDP); N-maleimidobenzyloxy-N-hydroxysuccinimide ester (MBS); N-maleimidobenzyloxy-N-hydroxysulfo succinimide ester (sulfo-MBS); succinimido 4-[p-maleimido dophenyl]butyrate (SMPB); sulfo succinimido 4-[p-maleimidophenyl]butyrate (sulfo-SMPB); N-[1,3’-Maleimidobutyloxy]succinimide ester (GBMS); N-[1,3’-maleimidobutyloxy]sulfo succinimide ester (sulfo-GBMS); N-[1,3’-maleimidocaproyloxy]succinimide ester (EMCS); N-[1,3’-maleimidocaproyloxy]sulfo succinimide ester (sulfo-EMCS); N-succinimidyl(4-iodoacetyl)aminobenzoate (SIAB); sulfo succinimidyl(4-iodosuccinimidyl)aminobenzoate (sulfo-SIAB); succinimidyl 4-[N-(maleimidomethyl) cyclohexane]-1-carboxylate (SMCC); sulfo succinimidyl 4-[N-(maleimidomethyl)cyclohexane]-1-carboxylate (sulfo-SMCC); succinimidyl 4-[N-(maleimidomethyl)cyclohexane]-1-carboxylate-(6-aminocaproate) (LC-SMCC); 4-succinimidoyloxyacryl methyl-2-pyridyldithio) tolune (SMPT); and sulfo-LC-SMPT.

[0516] As an exemplary protocol, a minicell suspension is made 5 mM EDTA/PBS, and a reducing solution of 2-mercaptoethanolimine in 5 mM EDTA/PBS is added to the minicells. The mixture is incubated for 90 minutes at 37°C. The minicells are washed with EDTA/PBS to remove excess 2-mercaptoethanolimine. The attachable moiety is dissolved in PBS, pH 7.2. A maleimide crosslinker is added to the solution, which is then incubated for 1 hour at room temperature. Excess maleimide is removed by column chromatography.

[0517] The minicells with reduced sulfhydryl groups are mixed with the derivatized compounds having an attachable moiety. The mixture is allowed to incubate at 4°C for 2 hours or overnight to allow maximum coupling. The conjugated minicells are washed to remove unreacted (unattached) comp-
pounds having the attachable moiety. Similar protocols are used for expressed membrane proteins with other reactive groups (e.g., carboxyl, amine) that can be conjugated to an attachable moiety.

[0518] IX.E. Non-Genetic Methods for Directing Compounds to Membranes

[0519] Included within the scope of the invention are compounds that can be inserted into the membrane of segregated micelles. Such compounds include attachable moieties that are chemically conjugated to the surface of a minicell, and compounds that associate with and/or insert into a membrane “spontaneously,” i.e., by virtue of their chemical nature. By way of non-limiting example, proteins that “spontaneously” insert into membranes include but are not limited to Thyka-loid membrane proteins (Woolhead et al., J. Biol. Chem. 276:14607-14613, 2001), the mitochondrial adenine nucleotide translocator (Jacot et al., J. Exp. Med. 193:509-519, 2001), and polypeptides obtained using the methods of Hunt et al. (Spontaneous, pH-dependent membrane insertion of a transbilayer alpha-helix, Biochem 36:15177-15192, 1997). Lipids, gangliosides, sphingomyelins, plasmalogens glyco- syl disaccharides, and sterols can also be incorporated into the membranes of segregated micelles.

[0520] X. Membrane Proteins

[0521] In certain aspects of the invention, membrane proteins from non-eubacterial organisms are expressed and displayed by micelles. The cellular membrane (a.k.a. the “plasma membrane”) is a lipid bilayer that forms the boundary between the interior of a cell and its external environment. The term “membrane proteins” refers to proteins that are found in membranes including without limitation cellular and organellar membranes.

[0522] X.A. Types of Membrane Proteins

[0523] X.A.1. In General

[0524] Membrane proteins consist, in general, of two types, peripheral membrane proteins and integral membrane proteins.

[0525] Integral membrane proteins can span both layers (or “leaflets”) of a lipid bilayer. Thus, such proteins may have extracellular, transmembrane, and intracellular domains. Extracellular domains are exposed to the external environment of the cell, whereas intracellular domains face the cytosol of the cell. The portion of an integral membrane protein that traverses the membrane is the “transmembrane domain.” Transmembrane domains traverse the cell membrane often by one or more regions comprising 15 to 25 hydrophobic amino acids which are predicted to adopt an alpha-helical conformation.

[0526] Integral membrane proteins are classified as bitopic or polytopic (Singer, 1990 Annu. Rev. Cell Biol. 6:247-96). Bitopic proteins span the membrane once while polytopic proteins contain multiple membrane-spanning segments.

[0527] A peripheral membrane protein is a membrane protein that is bound to the surface of the membrane and is not integrated into the hydrophobic layer of a membrane region. Peripheral membrane proteins do not span the membrane but instead are bound to the surface of a membrane, one layer of the lipid bilayer that forms a membrane, or the extracellular domain of an integral membrane protein.

[0528] X.A.2. In General

[0529] The invention can be applied to any membrane protein, including but not limited to the following exemplary receptors and membrane proteins. The proteins include but are not limited to are receptors (e.g., GPCRs, sphingolipid receptors, neurotransmitter receptors, sensory receptors, growth factor receptors, hormone receptors, chemokine receptors, cytokine receptors, immunological receptors, and compliment receptors, FC receptors), channels (e.g., potassium channels, sodium channels, calcium channels.), pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potassium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidase), enzymes and kinases (e.g., protein kinases, ATPases, GTPases, phosphatases, proteases), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN), chemotactic/adhesion proteins (e.g., ICAM1, selectins, CD34, VCAM-1, LFA-1, VLA-1), and phospholipases such as PI-specific PLC and other phospholipases.

[0530] X.A.3. Receptors

[0531] Within the scope of the invention are any receptor, including without limitation:

[0532] The nuclear receptors, e.g. the nuclear export receptor;

[0533] The peripheral (mitochondrial) benzodiazepine receptor (Gavish et al., “Enigma of the Peripheral Benzo-diazepine Receptor,” Pharmacological Reviews, Vol. 51, No. 4);

[0534] Adrenergic and muscarinic receptors (Brodde et al., “Adrenergic and Muscarinic Receptors in the Human Heart”, Pharmacological Review, Vol. 51, No. 4);

[0535] Gamma-aminobutyric acid (GABA) receptors (Barnard et al., “International Union of Pharmacology. IV. Subtypes of GABA Receptors: Classification on the Basis of Subunit Structure and Receptor Function,” Pharmacological Reviews, Vol. 50, No. 2);

[0536] Kinin B receptors (Marceau et al., “The B Receptors for Kinins,” Pharmacological Reviews, Vol. 50, No. 3);

[0537] Chemokine receptors (Murphy et al., “International Union of Pharmacology. XXII. Nomenclature for Chemokine Receptors” Pharmacological Reviews, Vol. 52, No. 1);

[0538] Glycine and NMDA Receptors (Dani et al., “Glycine and N-Methyl-D-Aspartate Receptors: Physiological Significance and Possible Therapeutic Applications,” Pharmacological Reviews, Vol. 50, No. 4);

[0539] Glutamate receptor ion channels (Dingledine et al., “The Glutamate Receptor Ion Channels”, Pharmacological Reviews, Vol. 51, No. 1);

[0540] Purine and pyrimidine receptors including purinergic (e.g., P2) receptors (Ralevic et al., “Receptors for Purines and Pyrimidines”, Pharmacological Reviews, Vol. 50, No. 3); CNS receptors and membrane transporters (E. Sylvester Vizi, “Role of High-Affinity Receptors and Membrane Transporters in Nonsynaptic Communication and Drug Action in the Central Nervous System,” Pharmacological Reviews, Vol. 52, No. 1);

[0541] Opioid receptors, including but not limited to the µ-opioid receptor (Quock et al., “The µ-Opioid Receptor: Molecular Pharmacology, Signal Transduction and the Determination of Drug Efficacy”, Pharmacological Review, Vol. 51, No. 3);

[0542] Angiotensin II receptors (Gasparo et al., “International Union of Pharmacology. XXIII. The Angiotensin II Receptors” Pharmacological Review, Vol. 52, No. 3);

[0543] Cholecystokinin receptors (Noble et al., “International Union of Pharmacology. XXI. Structure, Distribution,
and Functions of Cholecystokinin Receptors”, Pharmacological Reviews, Vol. 51, No. 4)

[0544] Hormone receptors, including but not limited to, the estrogen receptor; the glucocorticoid receptor; and the insulin receptor;

[0545] Receptors found predominantly in the central nervous system, including but not limited to, neuronal nicotinic acetylcholine receptors; the dopamine D2/D3 receptor; GABA receptors; central cannabinoid receptor CB1; opioid receptors, e.g., the kappa opioid receptor, and the methadone-specific opioid receptor; nicotinic acetylcholine receptors; serotonin receptors, e.g., the serotonin 5-HT3 receptor, the serotonin 5-HT4 receptor, and the serotonin-2 receptor; and dopamine receptors, e.g., the dopamine D2/D3 receptor; and the neurotensin receptor;

[0546] Receptors for growth factors, including but not limited to, the erythropoietin receptor; the FGF receptor; the VEGF receptor; VEGF receptor-2 protein; VEGF-receptor protein (KDR); fibroblast growth factor receptor; the p75 nerve growth factor receptor; epidermal growth factor receptor; IGFR-1 receptor; platelet factor-4 receptor; alpha platelet-derived growth factor receptor; hepatocyte growth factor receptor; and human fibroblast growth factor receptor;

[0547] Receptors for sphingolipids and lysophospholipids such as the Edg family of GPCRs;

[0548] Receptors for interleukins, e.g., receptors for interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, et seq.; and

[0549] Various receptors, including by way of non-limiting example, receptors described in U.S. Pat. No. 6,210,967 (DNA encoding a mammalian LPA receptor and uses thereof); U.S. Pat. No. 6,210,921 (CAR: a novel coxsackievirus and adenovirus receptor; U.S. Pat. No. 6,211,343 (Lactoferrin receptor protein; U.S. Pat. No. 6,218,509 (LH/CG receptor, DNA and uses thereof); U.S. Pat. No. 6,214,972 (DNA encoding prostaglandin receptor DP); U.S. Pat. No. 6,221,613 (DNA encoding a human melanin concentrating hormone receptor (MCH1) and uses thereof); U.S. Pat. No. 6,221,660 (DNA encoding SNORF2 receptor); U.S. Pat. No. 6,225,080 (Mu-subtype opioid receptor); U.S. Pat. No. 6,222,015 (Estrrogen receptor); U.S. Pat. No. 6,226,610 (Human metabotropic glutamate receptor subtypes (hmrk4, hmrR6, hmr7) and related DNA compounds); U.S. Pat. No. 6,255,496 (Nucleic acid encoding mammalian mu opioid receptor); U.S. Pat. No. 6,258,556 (cDNA and genomic clones encoding human mu opioid receptor and the purified gene product); U.S. Pat. No. 6,255,531 (Polynucleotide encoding insect ecdysone receptor); U.S. Pat. No. 6,255,531 (Glucan elicitor receptor, DNA molecule coding therefor, fungic-resistant plants transformed with the DNA molecule and method for creating the plants); U.S. Pat. No. 6,245,893 (Receptor that binds anti-convulsant compounds); U.S. Pat. No. 6,248,712 (Urokinase-type plasminogen activator receptor; U.S. Pat. No. 6,248,554 (DNA sequence coding for a BMP receptor); U.S. Pat. No. 6,248,520 (Nucleic acid molecules encoding nuclear hormone receptor coactivators and uses thereof); U.S. Pat. No. 6,242,251 (Rhesus neuroepithelial Y5 receptor); U.S. Pat. No. 6,252,056 (Human lysosphosphatidic acid receptor and use thereof); U.S. Pat. No. 6,255,472 (Isolated nucleic acid molecule encoding a human skeletal muscle-specific receptor); U.S. Pat. No. 6,291,207 (Herpes virus entry receptor protein); U.S. Pat. No. 6,291,206 (BMP receptor proteins); U.S. Pat. No. 6,291,195 (DNA encoding a human melanin concentrating hormone receptor (MCH1) and uses thereof); U.S. Pat. No. 6,344,200 (Lactoferrin receptor protein); U.S. Pat. No. 6,335,180 (Nucleic acid sequences encoding capsaicin receptor and uses thereof); U.S. Pat. No. 6,265,184 (Polynucleotides encoding chemokine receptor 8C); U.S. Pat. No. 6,207,799 (Neuropeptide Y receptor Y5 and nucleic acid sequences); U.S. Pat. No. 6,290,970 (Transferrin receptor protein of Moraxella); U.S. Pat. No. 6,326,350 (Transferrin receptor subunit proteins of Neisseria meningitidis); U.S. Pat. No. 6,313,279 (Human glutamate receptor and related DNA compounds); U.S. Pat. No. 6,313,276 (Human endothelin receptor); U.S. Pat. No. 6,307,030 (Androgen receptor proteins, recombinant DNA molecules coding for such, and use of such compositions); U.S. Pat. No. 6,306,622 (cDNA encoding a BMP type II receptor); U.S. Pat. No. 6,300,087 (DNA encoding a human serotonin receptor (5-HT4B) and uses thereof); U.S. Pat. No. 6,297,026 (Nucleic acids encoding the C140 receptor); U.S. Pat. No. 6,277,976 (Or-1, an orphan receptor belonging to the nuclear receptor family); U.S. Pat. No. 6,274,708 (Mouse interleukin-11 receptor); U.S. Pat. No. 6,271,347 (Eosinophil cation receptor); U.S. Pat. No. 6,262,016 (Transferrin receptor genes); U.S. Pat. No. 6,261,838 (Rat melanocortin receptor MC5-R); U.S. Pat. No. 6,258,943 (Human neurokinin-5 receptor); U.S. Pat. No. 6,284,878 (Gumma retinoic acid receptor); U.S. Pat. No. 6,258,944 (OB receptor isoforms and nucleic acids encoding them); U.S. Pat. No. 6,261,801 (Nucleic acids encoding tumor necrosis factor receptor 5); U.S. Pat. No. 6,261,800 (Luteinizing hormone/choriogonadotropin (LH/CG) receptor); U.S. Pat. No. 6,265,563 (Opioid receptor genes); U.S. Pat. No. 6,268,477 (Chemokine receptor 88-C); U.S. Pat. No. 6,316,611 (Human N-methyl-D-aspartate receptor subunits, nucleic acids encoding same and uses thereof); U.S. Pat. No. 6,316,604 (Human C3b/C4b receptor (CR1)); U.S. Pat. No. 6,287,855 (Nucleic acid encoding rat galanin receptor (GALR2)); U.S. Pat. No. 6,268,221 (Melanocyte stimulating hormone receptor and uses); and U.S. Pat. No. 6,268,214 (Vectors encoding a modified low affinity nerve growth factor receptor).

[0550] X.A.3. Other Membrane Proteins

[0551] Other membrane proteins are within the scope of the invention and include but are not limited to channels (e.g., potassium channels, sodium channels, calcium channels), pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potassium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidase), enzymes and kinases (e.g., protein kinases, ATPases, GTPases, phosphatases, proteases,), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN).


[0553] Cellular adhesion molecules, including but not limited to human rhinovirus receptor (ICAM-1), ICAM-2, ICAM-3, and PECAM-1, and chemotactic/adhesion proteins (e.g., selectins, CD34, VCAM-1, LFA-1, VLA-1) are within the scope of the invention. See also Alpin et al., "Signal Transduction and Signal Modulation by Cell Adhesion Receptors: The Role of Integrins, Cadherins, Immunoglobulin Cell Adhesion Molecules, and Selectins”, Pharmacological Reviews, Vol. 50, No. 2.

[0554] X.A.3.b. Cytochrome P450 Enzymes

[0555] The family of enzymes known as "cytochrome P450" enzymes (since they absorb light in the 450 nanometer
range), or as "cytochrome oxidase" enzymes (since they oxidize a wide range of compounds that do not naturally occur in circulating blood), are included within the scope of the invention. P450 enzymes encompass a variety of enzymes, many of which are involved in xenobiotic metabolism, including by way of non-limiting example the metabolism of drugs, prodrugs and toxins. Directories and databases of P450s, and information regarding their substrates, are available on-line (Fabian et al., The Directory of P450-containing Systems in 1996, Nucleic Acids Research 25:274-277, 1997). In humans, at least about 200 different P450s are present (for a review, see Hasler et al., Human cytochromes P450, Molecular Aspects of Medicine 20:1-137, 1999). There are multiple forms of these P450s and each of the individual forms exhibit degrees of specificity towards individual compounds or sets of compounds. In some cases, a substrate, whether it is a drug or a carcinogen, is metabolized by more than one cytochrome P450.

[0556] Members of the cytochrome P450 family are present in varying levels and their expression and activities are controlled by variables such as chemical environment, sex, developmental stage, nutrition and age. The cytochrome P450s are found at high concentrations in liver cells, and at lower concentrations in other organs and tissues such as the lungs (e.g., Forme-Pfister et al., Xenobiotic and endobiotic inhibitors of cytochrome P-450 dib function, the target of the debrisoquine/sparfloxacin type polymorphism, Biochem. Pharmacol. 37:3828-35, 1988). By oxidizing lipidophilic compounds, which makes them more water-soluble, cytochrome oxidase enzymes help the body eliminate (via urine, or in aerosols exhaled out of the lungs) compounds that might otherwise act as toxins or accumulate to undesired levels.

[0557] In humans, several cytochrome P450s have been identified as being involved in xenobiotic biotransformation. These include CYP1A1, CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP3A4, and CYP3A5 (Crespi et al., The use of heterologously expressed drug metabolizing enzymes—state of the art and prospects for the future, Pharm Ther 84:121-131, 1999).

[0558] X.A.Sc. Miscellaneous Membrane Proteins

[0559] In addition to the preceding non-limiting examples, the invention can be applied to the membrane proteins described in U.S. Pat. No. 6,335,018 (High molecular weight major outer membrane protein of moraxella); U.S. Pat. No. 6,264,954 (Haemophilus outer membrane protein); U.S. Pat. No. 6,197,543 (Human vesicle membrane protein-like proteins); U.S. Pat. No. 6,121,427 (Major outer membrane protein CD of branhamella); U.S. Pat. No. 6,089,743 and U.S. Pat. No. 6,013,514 (Haemophilus outer membrane protein); U.S. Pat. No. 6,004,562 (Outer membrane protein B1 of Moraxella catalarrhalis); U.S. Pat. No. 5,863,764 (DNA encoding a human membrane protein); U.S. Pat. No. 5,861,283 (DNA encoding a limbic system-associated membrane protein); U.S. Pat. No. 5,824,321 (Cloned leptospira outer membrane protein); U.S. Pat. No. 5,821,085 (Nucleotide sequences of a T. pallidum rare outer membrane protein); U.S. Pat. No. 5,821,055 (Chlamydia major outer membrane protein); U.S. Pat. No. 5,808,024 (Nucleic acids encoding high molecular weight major outer membrane protein of moraxella); U.S. Pat. No. 5,770,714 (Chlamydia major outer membrane protein); U.S. Pat. No. 5,763,589 (Human membrane protein); U.S. Pat. No. 5,753,459 (Nucleotide sequences of T. pallidum rare outer membrane protein); U.S. Pat. No. 5,607,920 (Concanavalin a binding proteins and a 76 kD chondrocyte membrane protein (CMP) from chondrocytes and methods for obtaining same); and U.S. Pat. No. 5,505,992 (DNA encoding the kD outer membrane protein of Haemophilus influenzae).

[0560] X.B. Membrane Anchoring Domains


[0562] X.C. Transmembrane Domains

[0563] A variety of types and examples of transmembrane domain are known. Proteins with up to 12 transmembrane domains are known (Fujiwara et al., Identification of thyroid hormone transporters in humans: different molecules are involved in a tissue-specific manner, Endocrinology 2001 142:2005-12; Sharina et al., Mutational analysis of the functional role of conserved arginine and lysine residues in transmembrane domains of the murine reduced folate carrier, Mol Pharmacol 2001 59:1022-8). However, the invention is not limited to any particular number of transmembrane domains.

[0564] Monotropic ("single pass") domains, which traverse a membrane once, include by way of non-limiting example, those found in receptors for epidermal growth factor (EGF), receptors for tumor necrosis factor (TNF) and the like. Polytropic ("multipass") proteins traverse a membrane two or more times. Non-limiting examples of polytropic proteins are as follows.

[0565] Biotropic ("2 passes") membrane proteins include, but are not limited to: Env2 of E. coli; the peroxisomal membrane protein Pex11p (Anton et al., ARF- and cuatomer-mediated peroxisomal vesiculation, Cell Biochem Biophys 2000; 32 Spring:27-36); pleiotropic drug ABC transporters of S. cerevisiae (Rogers et al., The pleiotropic drug ABC transporters from Saccharomyces cerevisiae, J Mol Microbiol Biotechnol 2001 3:207-14); and human and rat urine transporters hUAT and rUAT (Lipkowitz et al., Functional reconstitution, membrane targeting, genomic structure, and chromosomal localization of a human urine transporter, J Clin Invest 2001 107:1103-15).

[0566] Tritropic ("3 pass") membrane proteins include, but are not limited to: the ethylene receptor ETR1 of Arabidopsis; the Cauliflower Card Expression protein CCB1 (Palmer et al., A Brassica oleracea Gene Expressed in a Variety-Specific Manner May Encode a Novel Plant Transmembrane Receptor, Plant Cell Physiol 2001 42:404-413); and a splice variant of the mitochondrial membrane protein hMRPS3/4 (Li et al.,

Tetraspans or tetraspans are non-limiting examples of membrane proteins with four transmembrane domains. (Levy et al., J. Biol. Chem. 226:14597-14602, 1991; Tomlinson et al., J. Immunol. 23:136-40, 1993; and Barclay et al., (In) The Leucocyte antigen factbooks, Academic press, London, 1993). These proteins are collectively known as the ‘transmembrane 4 superfamily’ (TM4) because they span the plasma membrane four times. The proteins known to belong to this family include, but are not limited to: mammalian antigen CD9 (MC3), a protein involved in platelet activation and aggregation; mammalian leukocyte antigen CD37, expressed on B lymphocytes; mammalian leukocyte antigen CD53 (OX-44), which may be involved in growth regulation in hematopoietic cells; mammalian lysosomal membrane protein CD63 (Melanoma-associated antigen ME491; antigen A1); mammalian antigen CD81 (cell surface protein TAPA-1), which may play an important role in the regulation of lymphoma cell growth; mammalian antigen CD82 (Protein R2; Antibody C33; Kangai 1 (KAI1)), which associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway; mammalian antigen CD151 (SF-A1); Plateletendothelial tetraspan antigen 3 (PETA-3); mammalian TM4SF2 (Cell surface glycoprotein A15; TALLA-1; MXS1); mammalian TM4SF3 (Tumor-associated antigen CO-029); mammalian TM4SF6 (Tspan-6; TM4-D); mammalian TM4SF7 (Novel antigen 2 (NAG-2); Tspan-4); mammalian Tspan-2; mammalian Tspan-3 (TM4-A); mammalian tetraspan NET-5; and Schistosoma mansoni and japonicum 23 Kd surface antigen (SM23/SJ23).

Non-limiting examples of membrane proteins with six transmembrane domains include the EBV integral membrane protein LMP-1, and a splice variant of the mitochondrial protein hMRS3/4 (L et al., Characterization of a novel human putative mitochondrial transporter homologous to the yeast mitochondrial RNA splicing proteins 3 and 4, FEBS Lett 2001 Apr. 6; 494(1-2):79-84). Proteins with six transmembrane domains also include STEAP (six transmembrane epithelial antigens of the prostate) proteins (Aifar et al., U.S. Pat. No. 6,329,503). The prototype member of the STEAP family, STEAP-1, appears to be a type Ma membrane protein expressed predominantly in prostate cells in normal human tissues. Structurally, STEAP-1 is a 339 amino acid protein characterized by a molecular topology of six transmembrane domains and intracellular N- and C-termini, suggesting that it folds in a “serpentine” manner into three extracellular and two intracellular loops.

Literally hundreds of 7-pass membrane proteins are known. G-protein coupled receptors (GPCRs), including without limitation beta-adreno receptors, adrenergic receptors, EDG receptors, adenosine receptors, B receptors for kinins, angiotensin receptors, and opioid receptors are of particular interest. GPCRs are described in more detail elsewhere herein.

A non-limiting example of a protein with 9 transmembrane domains is Lipocalin-1 interacting membrane receptor (Wojnar et al., Molecular cloning of a novel Lipocalin-1 interacting human cell membrane receptor (LIMR) using phage-display, J Biol Chem 2001 3; [epub ahead of print]).

Proteins with both transmembrane and anchoring domains are known. For example, AMPA receptor subunits have transmembrane domains and one membrane-anchoring domain.

A variety of databases that describe known, and software programs that predict, membrane anchoring and transmembrane domains are available to those skilled in the art. See, for example Gerd.bdb GCRDb [G Protein Coupled Receptor database], Tmbase.bdb Tmbase [databases of transmembrane domains], Predom.srv Propom [Protein domains], Tmmap.srv TMAP [Protein transmembrane segments prediction], Tm7.srv TM7 [Retrieval of data on G protein-coupled receptors], and Memsat.sof MEMSAT [transmembrane structure prediction program].

Quentin and Fichant (J Mol Microbiol Biotechnol 2000 2:501-4, ABCdb: an ABC transporter database) have described a database devoted to the ATP-binding cassette (ABC) protein domains (ABCdb), the majority of which energize the transport of compounds across membranes. In bacteria, ABC transporters are involved in the uptake of a wide range of molecules and in mechanisms of virulence and antibiotic resistance. In eukaryotes, most ABC transporters are involved in drug resistance, and many are associated with diseases. ABCdb can be accessed via the World Wide Web (http://ii1cb.canr.msu.edu/ABCdb/). See also Sanchez-Fernandez et al., The Arabidopsis thaliana ABC protein superfamily: a complete inventory, J Biol Chem 2001 May 9; [epub ahead of print], and Rogers et al., The pleitropic drug ABC transporters from Saccharomyces cerevisiae, J Mol Microbiol Biotechnol 2001 April 3(2):207-14.

X.D. Functions and Activities of Membrane Proteins

Non-limiting examples of membrane proteins include membrane-associated enzymes. Membrane-associated enzymes include but not limited to certain enzymes of the electron transport chain (ETC), antigenic proteins such as the major histocompatibility (MHC) antigens, transport proteins, channels, hormone receptors, cytokine receptors, glucose permeases, gap junction proteins and bacteriophodopsins.

A “transport protein” or “transporter” is a type of membrane protein that allows substances to cross plasma membranes at a rate that is faster than what is found by diffusion alone. Some transport proteins expend energy to move substances (active transport). Many active transport proteins are ATPases (e.g., the Na+-K+ ATPase), or at least bind ATP by virtue of comprising an ATP-binding cassette (ABC) (see, e.g., Rogers et al., The pleitropic drug ABC transporters from Saccharomyces cerevisiae, J Mol Microbiol Biotechnol 3:207-14, 2001). Nucleobase transporters are reviewed by De Konings and Dallinas (Nucleobase Transporters, Mol Membr Biol 17:75-94, 2000).

A “channel protein” is a protein that facilitates the diffusion of molecules/ions across lipid membranes by forming a hydrophilic pore or “channel” that provides molecules/ions access through lipid membranes, which are generally hydrophobic. Channels are often multimeric, with the pore being formed by subunit-subunit interactions.

A “receptor” is a molecular entity, typically a protein, that is displayed on the surface of a cell. A receptor is characterized by high affinity, often a specific binding of a specific substance, typically resulting in a specific biochemical or physiological effect.

A “hormone” is a naturally occurring substance secreted by specialized cells that affects the metabolism or...
behavior of other cells having receptors for the hormone. Non-limiting examples of hormones having include but are not limited to insulin, cytokines, steroid hormones, histamines, glucagon, angiotensin, catecholamines, low density lipids (LDLs), tumor necrosis factor alpha, tumor necrosis factor beta, estrogen, and testosterone.

[0580] X.E. G-Protein-Coupled Receptors

[0581] G protein-coupled receptors (GPCRs) constitute the most prominent family of validated drug targets within biomedical research and are thought to be involved in such diseases and disorders as heart disease, hypertension, cancer, obesity, and depression and other mental illnesses. Over half of approved drugs elicit their therapeutic effects by selectively addressing members of this target family and more than 1000 sequences of the human genome encode for GPCRs containing the classical 7-pass membrane structure characteristic of this family of proteins (Marinissen, M. and J. S. Gutkind, G-protein-coupled receptors and signaling networks: emerging paradigms (Review), Trends Pharmacol. Sci. 22: 368-376, 2001). Many pharmacological drug companies are interested in the study of G-coupled proteins. It is possible to co-express a G-coupled protein receptor and its associated G-protein to study their pharmacological characteristics (Strosberg and Marullo, Functional expression of receptors in microorganisms. TIPS, 1992: 13: 95-98).


[0584] GPCRs belong to a superfamily of at least 6 families of receptors, the most important of which is the main family, A. Members of the membrane protein gene superfamily of GPCRs have been characterized as having seven putative transmembrane domains. The transmembrane domains are believed to represent transmembrane alpha-helices connected by extracellular or cytoplasmic loops. A functional G-protein is a trimer which consists of a variable alpha subunit coupled to much more tightly-associated and constant beta and gamma subunits, although G-protein independent actions have been postulated (Marinissen, M. and J. S. Gutkind, G-protein-coupled receptors and signaling networks: emerging paradigms, Trends. Pharmacol. Sci. 22: 368-376, 2001 Review). A variety of ligands have been identified which function through GPCRs. In general, binding of an appropriate ligand (e.g., bioactive lipids, ions, bioactive amines, photons, odorants, hormones, neurotransmitters, peptides, nucleosides, etc.) to a GPCR leads to the activation of the receptor. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors. Typically, activation of a GPCR initiates the regulatory cycle of a corresponding G-protein. This cycle consists of GTP exchange for GDP, dissociation of the alpha and beta/gamma subunits, activation of the second messenger pathway by a complex of GTP and the alpha subunit of the G-protein, and return to the resting state by GTP hydrolysis via the innate GTPase activity of the G-protein alpha subunit A.

[0585] GPCRs include, without limitation, dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological disorders. Other examples of members of this family include caeruloplasmin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins and rhodopsins, odorant, cytomembranous receptors, and the like.

[0586] Most GPCRs have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The seven transmembrane regions, each comprising conserved hydrophobic stretches of about 20 to 30 amino acids, are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 is also implicated in signal transduction.

[0587] Although not wishing to be bound by any particular theory, it is believed that GPCRs participate in cell signaling through their interactions with heterotrimeric G-proteins composed of alpha, beta and gamma subunits (Marinissen, M. and J. S. Gutkind, G-protein-coupled receptors and signaling networks: emerging paradigms, Trends. Pharmacol. Sci. 22:368-376, 2001). In some aspects of the invention, GPCRs and homologs are displayed on the surfaces of minicells.

[0588] X.F. EDG Receptors and Other Sphingolipid-Binding Receptors

[0589] The Endothelial Differentiation Gene (EDG) receptor family includes but is not limited to eight presently known GPCRs that have a high affinity to lipid ligands (Lynch et al., Life on the edg. Trends Pharmacol. Sci., 1999. 20: 273-5). These transmembrane receptors are found in several different tissues in different species. EDG receptors have been shown to be involved in calcium mobilization, activation of mitogen-activated protein kinase, inhibition of adenylate cyclase activation, and alterations of the cytoskeleton. The EDG family is divided into two different groups based on homology and ligand specificity. The EDG 2, 4, and 7 receptors are specific for the ligand lysophosphatidic acid (LPA) (An et al., Signaling Mechanism and molecular characteristics of G protein-coupled receptors for lysophosphatidic acid and sphingosine 1-phosphate. J Cell Biochem, 30/31:147-157, 1998; Goetzl

Receptors that bind SIP and other sphingolipids are used in one aspect of the invention (for a review of some SIP-binding receptors, see Spiegel et al., Biochem. Biophys. Acta 1484:107-116, 2000). Such receptors include but are not limited to members of the EDG family of receptors (a.k.a. 1p1 receptors, Chun, Crit. Rev. Neuro. 13:151-168, 1999), and isoforms and homologs thereof such as NRG1 and AGR16.

EDG-1 was the first identified member of a class of G protein-coupled endothelial-derived receptors (EDGs). Non-limiting examples of other EDG family members that also bind SIP include EDG-3 (a.k.a. AGR16; the rat homolog of EDG-3 is designated H218), EDG-5, EDG-6 and EDG-8. For reviews, see Goetzl et al., Adv. Exp. Med. Biol. 469:259-264, 1999; and Chun et al., Cell. Biochem. Biophys. 30:213-242, 1999).


EDG-6 is described by Graler et al. (Genomics 53:164-169, 1998), Yamazaki et al. (Biochem. Biophys. Res. Commun. 268:583-589, 2000), and Van Brocklyn et al. (Sphingosine-1-phosphate is a ligand for the G protein-coupled receptor EDG-6, Blood 95:2624-9, 2000).

EDG-8 from rat brain is described by Im et al., (J. Biol. Chem. 275:14281-14286, 2000). Homologs of EDG-8 from other species, including humans, may also be used in the present invention.

The Mil receptor (Mil is an abbreviation for "miles apart") binds SIT and regulates cell migration during vertebrate heart development. The Mil receptor of Zebrafish is described by Mohler et al. (J. Immunol. 151:1548-1561, 1993). Another SIP receptor is NRG1 (nerve growth factor regulated gene-1), the rat version of which has been identified (Glickman et al., Mol. Cel. Neurosci. 14:141-152, 1999).

Receptors that bind sphingosylphosphoryl cholines (SPC) are also used in this aspect of the invention. Such receptors include but are not limited to members of the SCA,MPTER family of receptors (Mao et al., Proc. Natl. Acad. Sci. U.S.A. 93:1993-1996, 1996; Betto et al., Biochem. J. 322:327-333, 1997). Some evidence suggests that EDG-3 may bind SPC in addition to SIP (Okamoto et al., Biochem. Biophys. Res. Commun. 260:203-208, 1999). Derivatives of EDG-3 that bind both SIT and SPC are used in one aspect of the invention.


XI. Recombinant DNA Expression

In order to achieve recombinant expression of a fusion protein, an expression cassette or construct capable of expressing a chimeric reading frame is introduced into an appropriate host cell to generate an expression system. The expression cassettes and constructs of the invention may be introduced into a recipient eubacterial or eukaryotic cell either as a nonreplicating DNA or RNA molecule, which may be a linear molecule or, more preferably, a closed covalent circular molecule. Since such molecules are incapable of autonomous replication, the expression of the gene may occur through the transient expression of the introduced sequence. Alternatively, permanent expression may occur through the integration of the introduced DNA sequence into the host chromosome.

XI.A. Recombinant DNA Expression Systems

A variety of eubacterial recombinant DNA expression systems may be used to produce the fusion proteins of the invention. Host cells that may be used in the expression systems of the present invention are not strictly limited, provided that they are suitable for use in the expression of the fusion protein of interest and can produce mictelines. Non-limiting examples of recognized eubacterial hosts that may be used in
the present invention include bacteria such as E. coli, Bacillus, Streptomyces, Pseudomonas, Salmonella, Serratia, and the like. [0605] Eubacterial expression systems utilize plasmid and viral (bacteriophage) expression vectors that contain replication sites and control sequences derived from a species compatible with the host may be used. Suitable plagues or bacteriophage vectors include E. coli, B. subtilis, and the like. Suitable virus vectors may include pMAM-neo, pRK1 and the like. Appropriate eubacterial plasmid vectors include those capable of replication in E. coli (such as, by way of non-limiting example, pBR322, pUC118, pUC119, ColEl, pSC101, pACYC184, OXY. See “Molecular Cloning: A Laboratory Manual” (1989) Bacillus plasmids include pC194, pC221, pT127, and the like (Gryczan. In: The Molecular Biology of the Bacillus, Academic Press, NY, pp. 307-329, 1982). Suitable Streptomyces plasmids include pJ101 (Kendall et al., J. Bacteriol. 169:4777-4783, 1987), and Streptomyces bacteriophages such as C3L (Chater et al., In: Sixth International Symposium on Actinomycetes Biology, Akademiai Kaido, Budapest, Hungary, pp. 45-54, 1986). Pseudomonas plasmids are reviewed by John et al. (Rev. Infect. Dis. 8:693-704, 1986), and Izaki (Jpn. J. Bacteriol. 33:729-742, 1978). See also Brent et al., “Vectors Derived From Plasmids,” Section II, and Lech et al. “Vectors derived from Lambda and Related Bacteriophages” Section III, in Chapter 1 of Short Protocols in Molecular Biology, 2nd Ed., Ausubel et al., eds., John Wiley and Sons, New York, 1992, pages 1-13 to 1-27; Lech et al., “Vectors derived from Lambda and Related Bacteriophages” Section III and Id. pages 1-28 to page 1-52. [0606] To express a protein, including but not limited to a fusion protein, in a eubacterial cell, it is necessary to operably link the ORF encoding the protein to a functional eubacterial or viral promoter. Such promoters may be either constitutive or, more preferably, regulatable (i.e., inducible or derepressible). Examples of constitutive promoters include the int promoter of bacteriophage lambda, the bla promoter of the lactamase gene sequence of pBR322, and the cat promoter of the chloramphenicol acetyl transferase gene sequence of pBR325, and the like. Examples of inducible eubacterial promoters include the major right and left promoters of bacteriophage lambda (P and P), the trp, recA, lacZ, lac, and gal promoters of E. coli, the alpha-amylase (Ullman et al., J. Bacteriol. 162:176-182, 1985) and the sigma-28-specific promoters of B. subtilis (Gilman et al., Gene Sequence 32:11-20, 1984), the promoters of the bacteriophages of Bacillus (Gryczan, in: The Molecular Biology of the Bacilli, Academic Press, Inc., NY, 1982), and Streptomyces promoters (Ward et al., Mol. Gen. Genet. 203:468-478, 1986). Eubacterial promoters are reviewed by Glick (Ind. Microbiol. 1:277-282, 1987), Cenatiempo (Biochimie 68:505-516, 1986), and Gottesman (Ann. Rev. Genet. 18:415-442, 1984). [0607] Proper expression also requires the presence of a ribosome-binding site upstream of the gene sequence-encoding sequence. Such ribosome-binding sites are disclosed, for example, by Gold et al. (Ann. Rev. Microbiol. 35:365-404, 1981). The selection of control sequences, expression vectors, transformation methods, and the like, are dependent on the type of host cell used to express the gene. As used herein, “cell”, “cell line”, and “cell culture” may be used interchangeably and all such designations include progeny. Thus, the words “transformants” or “transformed cells” include the primary subject cell and cultures derived therefrom, without regard to the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. However, as defined, mutant progeny have the same functionality as that of the originally transformed cell. [0608] Mammalian expression systems utilize host cells such as HeLa cells, cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin and their derivatives. Preferred mammalian host cells include SP2/0 and J558L, as well as neuroblastoma cell lines such as IMR 322, which may provide better capacities for correct post-translational processing. Non-limiting examples of mammalian extrachromosomal expression vectors include pCR3.1 and pCDNA3.1, and derivatives thereof including but not limited to those that are described by and are commercially available from Invitrogen (Carlsbad, Calif.). [0609] Several expression vectors are available for the expression of polypeptides in mammalian host cells. A wide variety of transcriptional and translational regulatory sequences may be employed, depending upon the nature of the host. The transcriptional and translational regulatory signals may be derived from viral sources, such as adenovirus, bovine papilloma virus, cytomegalovirus (CMV), simian virus, or the like, where the regulatory signals are associated with a particular gene sequence which has a high level of expression. Alternatively, promoters from mammalian expression products, such as actin, collagen, myosin, and the like, may be employed. Transcriptional initiation regulatory signals may be selected which allow for repression or activation, so that expression of the gene sequences can be modulated. Of interest are regulatory signals that are temperature-sensitive since, by varying the temperature, expression can be repressed or initiated, or are subject to chemical (such as metabolite) regulation. [0610] Preferred eukaryotic plasmids include, for example, BPV, vaccinia, SV40, 2-micron circle, and the like, or their derivatives. Such plasmids are well known in the art (Botstein et al., Miami Wmr. Symp. 19:265-274, 1982; Broach, in: The Molecular Biology of the Yeast S. cerevisiae, Life Cycle and Inheritance, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., p. 445-470, 1981; Broach, Cell 28:203-204, 1982; Bollon et al., J. Clin. Hematol. Oncol. 10:39-48, 1980; Manniti, In: Cell Biology A Comprehensive Treatise, Vol. 3, Gene Sequence Expression, Academic Press, NY, pp. 563-608, 1980). [0611] Expression of polypeptides in eukaryotic hosts generally involves the use of eukaryotic regulatory regions. Such regions will, in general, include a promoter region sufficient to direct the initiation of RNA synthesis. Preferred eukaryotic promoters include, for example, the promoter of the mouse metallothionein I gene sequence (Hamer et al., J. Mol. Appl. Gen. 1:273-288, 1982); the TK promoter of Herpes virus (McKnight, Cell 31:355-365, 1982); the SV40 early promoter (Benoi et al., Nature (London) 290:304-31, 1981); and the yeast gal4 gene sequence promoter (Johnston et al., Proc. Natl. Acad. Sci. (USA) 79:6971-6975, 1982, Silver et al., Proc. Natl. Acad. Sci. (USA) 81:5951-5955, 1984). [0612] Expression sequences and elements are also required for efficient expression. Non-limiting examples include Kozak and IRES elements in eukaryotes, and Shine-Delgarno sequences in prokaryotes, which direct the initiation of translation (Kozak, Initiation of Translation in Prokaryotes and Eukaryotes, Gene, 1999, 234: 187-208; Martinez-Salas et al., Functional interactions in internal translation
initiation directed by viral and cellular IRES elements. Jour. of Gen. Virol. 82:973-984, 2001); enhancer sequences; optional sites for repressor and inducers to bind; and recognition sites for enzymes that cleave DNA or RNA in a site-specific manner. Translation of mRNA is generally initiated at the codon which encodes the first methionine; if so, it is preferable to ensure that the linkage between a eukaryotic promoter and a preselected ORF does not contain any intervening codons that encode a methionine (i.e., AUG). The presence of such codons results either in the formation of a fusion protein with an uncharacterized N-terminal extension (if the AUG codon is in the same reading frame as the ORF) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the ORF).

[0613] XLI.B. Expression of Membrane Proteins Presently, the most commonly used expression systems for the expression of integral membrane proteins are eukaryotic and eubacterial whole cell expression systems. Although minicells have been used to express several eubacterial membrane proteins, the production of non-eubacterial membrane proteins has not been reported. One aspect of the invention is the discovery that the minicell expression system can be made to express and preferably display integral membrane proteins from non-eubacterial organisms.

[0614] Some commonly used expression systems include in vitro systems, such as the Rabbit Retinococyte Lysozyme System and E. coli S30 Extract System (both available from Promega) (Zubay, Methods Enz. 65:856, 1980) and in vivo systems, such as eukaryotic cell culture expression, and bacterial expression systems. Although this is not an exhaustive list, these systems are representative.

[0615] The Rabbit Retinococyte Lysozyme system utilizes a cell lysate that contains all the enzymes required for transcription and translation to drive protein expression, and is a good in vitro system for producing small amounts of labeled and unlabeled protein. However, this system is not well-suited for the production of large quantities of proteins and is limited to soluble proteins as there are no membranes in which to incorporate membrane proteins.

[0616] In eukaryotic cell culture systems, expression vectors suited for expression in host eukaryotic cells are transfected into cultured cells and protein is translated from mRNA produced from the vector DNA template Kaufman, Overview of vector design for mammalian gene expression. Mol Biotechnol, 2001. 16: 151-160; Lee et al., Heterologous gene expression in avian cells: Potential as a producer of recombinant proteins. J Biomed Sci, 1999. 6: 8-17; Voorma et al., Initiation of protein synthesis in eukaryotes. Mol Biol Rep, 1994. 19: 139-45). Cells can then either be harvested to prepare at least partially purified proteins or proteins produced from the expression element can be studied in the host cell environment.

[0617] Regarding membrane proteins, such systems have limitations. Primary cell lines are difficult to maintain and are short lived. Immortalized cell lines divide indefinitely, but have been altered in many ways and can be unpredictable. The transfection efficiency is very low in most eukaryotic cells and some cell types are refractory to transformation. Moreover, other proteins are expressed in these cells along with the protein of interest. This can cause difficulties when performing certain experiments and when attempting to immunoprecipitate the protein. Good experimental data are difficult to obtain from studies such as binding assays (because of high background due to endogenous proteins), and crystal determination of protein structure (because it is difficult to obtain enough purified protein to efficiently form crystals).


[0619] In bacterial expression systems, bacterial cells are transformed with expression elements, and transcription and translation is driven from a bacterial promoter. Bacteria divide very rapidly and are easy to culture; it is relatively easy to produce a large number of bacteria in a short time. Moreover, incorporation of expression elements vector into bacterial cells is efficient. Transformed cells can be isolated that arise from a single bacterium. Cultures of transformed cells are thus genetically identical and all cells in the culture will contain the expression element. However, there are proteins that are not suitable for expression in bacteria because of differences between eukaryotic cells and bacterial cells in transcription, translation, and post-translational modification.

[0620] The E. coli whole cell expression system has been used to express functional integral membrane proteins. For a review, see Strosberg, Functional expression of receptors in microorganisms. TiPS, 1992. 13: 95-98. Examples of mammalian integral membrane proteins that have been expressed in Escherichia coli include rat alphab-2B-adrenoceptors (Xia et al., Functional expression of rat β2-adrenoceptor in E. coli. Euro J. Pharma, 1993. 246: 129-133) and the human beta2-adrenergic receptor (Marullo et al., Human β 2-adrenergic receptors expressed in Escherichia coli membranes retain their pharmacological properties. Proc. Natl. Acad. Sci. USA, 1988. 85: 7551-7555). In some of these studies, the integral membrane proteins were not only expressed in E. coli expression systems, but also retained their pharmacological properties. This allows for binding studies to be performed with minimal background signal (“noise”) from host cell proteins. It has also been shown that signal sequences (the short hydrophobic amino acid sequence at the N-terminus of integral membrane proteins that signals the transport of the protein to the membrane) from mammalian cells may be functional in the E. coli system.

[0621] As is discussed herein, the expression of membrane proteins such as GPCRs, ion channels, and immuno-receptors in minicells, and their incorporation into the membranes thereof, allows for the study and use of such non-eubacterial membrane proteins. The minicell system of the invention is particularly well-suited for the study and expression of EDG proteins because of the lipid nature of the ligands for these receptors. The identification of ligand binding kinetics and biochemistry of these receptors because of the physiochemical properties of the lipid ligands (LPA and SIP), which result in high non-specific binding (Lee et al., Sphingosine-1-phosphate is a ligand for the G protein-coupled receptor EDG-1. Science, 1998. 270: 1552-1555; Van Brocklyn et al., Sphingosine-1-phosphate is a ligand for the G protein-coupled receptor EDG-6. Blood, 2000. 95: 2624-2629; Liu et

[0622] It is believed, for example, that in the case of the ion channels, the minicell expression system is less cumbersome than procedures that are presently used to study properties of ion channels, such as, e.g., reconstitution studies (Montal, Molecular anatomy and molecular design of channel proteins. FASEB J., 1990. 4: p. 2623-2635). Ionic conditions both inside and outside of minicells can be manipulated in various ways, and the properties of an ion channel that is expressed in a minicell, and factors that activate or modulate the activities of the channel, can be studied. Binding and kinetic studies are performed on ligand-mediated ion channels. This type of study is enhanced when the ion channel is able to interact specifically with its ligand and has a low background of non-specific binding from the endogenous proteins. This can be accomplished by making the minicells into protoplasts or poroplasts in which the ligand-activated ion channels in the inner membrane are exposed to the external environment and have better access to their specific ligand.

[0623] A “recombinant expression system” (or simply “expression system”) is one that directs the production of exogenous gene products in a host cell or minicell of choice. By “expressed” it is meant that a gene product of interest (which can be a protein or nucleic acid) is produced in the expression system of choice.

[0624] Host cells (and/or minicells) harboring an expression construct are components of expression systems. An “expression vector” is an artificial nucleic acid molecule into which an exogenous ORF encoding a protein, or a template of a bioactive nucleic acid can be inserted in such a manner so as to be operably linked to appropriate expression sequences that direct the expression of the exogenous gene. By the term “operably linked” it is meant that the part of a gene that is transcribed is correctly aligned and positioned with respect to expression sequences that promote, are needed for and/or regulate this transcription. The term “gene product” refers to either a nucleic acid (the product of transcription, reverse transcription, or replication) or a polypeptide (the product of translation) that is produced using the non-vector nucleic acid sequences as a template.

[0625] In some applications, it is preferable to use an expression construct that is an episomal element. If the episomal expression construct expresses (or, preferably in some applications, over-expresses) a an ORF that has been incorporated into the episomal expression construct, the minicells will direct the production of the polypeptide encoded by the ORF. At the same time, any mRNA molecules transcribed from a chromosomal gene prior to minicell formation that have been transferred to the minicell are degraded by endogenous RNases without being replaced by new transcription from the (absent) bacterial chromosome.

[0626] Chromosomally-encoded mRNAs will not be produced in minicells and will be “diluted” as increasing amounts of mRNAs transcribed from the episomal element are generated. A similar dilution effect is expected to increase the relative amount of episomal-generated proteins relative to any chromosomally-encoded proteins present in the minicells. It is thus possible to generate minicells that are enriched for proteins encoded by and expressed from episomal expression constructs.

[0627] Although by no means exhaustive, a list of episomal expression vectors that have been expressed in eubacterial minicells is presented in Table 4.

[0628] It is also possible to transform minicells with exogenous DNA after they have been prepared or separated from their parent cells. For example, phage RNA is produced in minicells after infection by lambda phage (Witkiewicz and Taylor, Ribonucleic acid synthesis after adsorption of the bacteriophage lambda on Escherichia coli minicells, Acta Microbiol Pol A 7:21-4, 1975), even though replication of lambda phage may not occur in minicells (Witkiewicz and Taylor, The fate of phage lambda DNA in lambda-infected minicells, Biochim Biophys Acta 564:31-6, 1979).

[0629] Because it is the most characterized minicell-producing species, many of these episomal elements have been examined in minicells derived from E. coli. It is understood by practitioners of the art, however, that many episomal elements that are expressed in E. coli also function in other eubacterial species, and that episomal expression elements for minicell systems in other species are available for use in the invention disclosed herein.

[0630] In one aspect of the invention, eukaryotic and archaebacterial minicells are used for expression of membrane proteins, particularly in instances where such desirable proteins have enhanced or altered activity after they undergo post-translational modification processes such as phosphorylation, proteolysis, myristilation, GPI anchoring and glycosylation. Expression elements comprising expression sequence operably linked to ORF's encoding the membrane proteins of interest are transformed into eukaryotic cells according to methods and using expression vectors known in the art. By way of non-limiting example, primary cultures of rat cardiomyocytes have been used to produce exogenous proteins after transfection of expression elements therefor by electroporation (Nakajima et al., Expression and characterization of Edg-1 receptors in rat cardiomyocytes: Calcium deregulation in response to sphingosine-1-phosphate, Eur. J. Biochem. 267: 5679-5686, 2000).

[0631] Yeast cells that produce minicells are transformed with expression elements comprising an ORF encoding a membrane protein operably linked to yeast expression sequences. Cells that harbor a transformed expression element may be selected using a gene that is part of the expression element that confers resistant to an antibiotic, e.g., neomycin.

[0632] Alternatively, in one aspect of the invention, bacterial minicells are prepared that contain expression elements that are prepared from shuttle vectors. A “shuttle vector” has sequences required for its replication and maintenance in cells from two different species of organisms, as well as expression elements, at least one of which is functional in bacterial cells, and at least one of which is functional in yeast cells. For example, E. coli-yeast shuttle vectors are known in the art and include, by way of non-limiting example, those derived from Yip, Yrp, Yep and Yep. Preferred E. coli-yeast shuttle vectors are episomal elements that can segregate into yeast minicells (i.e., Yrp, Yep and Yep. Particularly preferred are expression vectors of the Yep (yeast episomal plasmid) class, and other derivatives of the naturally occurring yeast plasmid known as the 2 μm circle. The latter vectors have relatively high transformation frequencies and are stably maintained through mitosis and meiosis in high copy number.
<table>
<thead>
<tr>
<th>EPISOMAL ELEMENT</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR1</td>
<td>Hochmunova et al., <em>Folia Microbiol.</em> (Praga) 26: 270-276</td>
</tr>
<tr>
<td>R6H</td>
<td>Hochmunova et al., <em>Folia Microbiol.</em> (Praga) 26: 270-276</td>
</tr>
<tr>
<td>PGR2.1</td>
<td>Rigg et al., <em>Arch. Oral. Biol.</em> 45: 41-52 (2000); expresses cell surface antigen of <em>E. gingivalis</em></td>
</tr>
</tbody>
</table>

"mini-plasmid" derivative of RK2

CoEl1  

PSC101  
Rastkchain et al., *J. Bacteriol.* 165: 82-87 (1986); Curtiss, Roy, III, U.S. Pat. No. 4,190,495; Issued Feb. 26, 1980 |

pACYC184  

ColElb, ColElb DRD & pUC19  

R-plasmid  

PCRI  
Hollenberg et al., *Gene* 1: 33-47 (1976); yeast shuttle vector |

Bacteriophage

Lambda  

M13  

T7  

P1  
Curtiss, Roy, III, U.S. Pat. No. 4,190,495; Issued Feb. 26, 1980;  

[0633] For expression of membrane proteins, and/or other proteins of interest in the recipient cell, ORFs encoding such proteins are operably linked to eukaryotic expression sequences that are appropriate for the recipient cell. For example, in the case of *E. coli*-yeast shuttle vectors, the ORFs are operably linked to expression sequences that function in yeast cells and/or minicells. In order to assess the effectiveness of a gene delivery vehicle, or a gene therapy expression element, an ORF encoding a detectable polypeptide (e.g., GFP, beta-galactosidase) is used. Because the detectable polypeptide is operably linked to eukaryotic expression elements, it is not expressed unless it has been transferred to its recipient (eukaryotic) cell. The signal from the detectable polypeptide thus correlates with the efficiency of gene transfer by a gene delivery agent, or the degree of expression of a eukaryotic expression element.

[0634] Gyris and Duda (High-efficiency transformation of *Saccharomyces* cells by bacterial minicell protoplast fusion, *Mol. Cell Biol.* 6:329507, 1986) allegedly demonstrated the transfer of plasmid molecular by fusing minicell protoplasts with yeast protoplasts. Gyris and Duda state that 10% of *Saccharomyces cerevisiae* cells were found to contain transforming DNA sequences. However, the plasmids did not contain eukaryotic expression elements, were not shuttle vectors, and genetic expression of the plasmids in yeast cells was not examined.

[0635] X. Uses of Minicells in Research

[0636] XII A. In General

[0637] The minicells of the invention can be used in research applications such as, by way of non-limiting example, proteomics, physiology, chemistry, molecular biology, physics, genetics, immunology, microbiology, proteomics, virology, pathology, botany, and neurobiology. Research applications include but are not limited to protein-ligand binding studies, competitive inhibition studies, structural studies, protein interaction studies, transfection, signaling studies, viral interaction studies, ELISA, antibody studies, gel electrophoresis, nucleotide acid) applications, peptide production, cell culture applications, cell transport studies, isolation and separation studies, chromatography, labeling studies, synthesis of chemicals, chemical cross linking, flow cytometry, nanotechnology, micro switches, micro-machines, and agricultural studies, cell death studies, cell-cell interactions, proliferation studies, and protein-drug interactions. Minicells are applicable to research applications involving, by way of non-limiting example, the elucidation, manipulation, production, replication, structure, modeling, observations, and characterization of proteins.

[0638] The types of proteins that can be involved in research applications of minicells can be either soluble proteins or membrane bound proteins, and include but are not limited to receptors (e.g., GPCRs, sphingolipid receptors, neurotransmitter receptors, sensory receptors, growth factor receptors, hormone receptors, chemokine receptors, cytokine receptors, immunological receptors, and compliment receptors, FC receptors), channels (e.g., potassium channels, sodium channels, calcium channels, pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potas
sium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidases), enzymes and kinases (e.g., protein kinases, ATPases, GTPases, phosphatases, proteases), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN), chemotactic/adhesion proteins (e.g., ICAM1, selectins, CD34, VCAM-1, LFA-1, VLA-1), and chimeric/fusion proteins (e.g., proteins in which a normally soluble protein is attached to a transmembrane region of another protein).

[0639] Research products are designed for any specific type of application. These products may be packaged and distributed as, by way of non-limiting example, kits, chemicals, solutions, buffers, powders, solids, filters, columns, gels, matrices, emulsions, pellets, capsules, and aerosols. Kits and reagents for certain research applications may be required by regulatory agency to be labeled “research use only” in order to indicate that the reagents are not intended for use in humans.

[0640] XII.B. Transfection

[0641] Transfection is the process of introducing genetic material into eukaryotic and archaeabacterial cells using biological, biochemical or physical methods. This process allows researchers to express and study target proteins in cultured cells (research use) as well as to deliver genetic material to cells in vivo or ex vivo systems (gene therapy). There are a variety of techniques which allow for the introduction and expression of proteins into target cells. These include mechanical transfection (Biolistic particles and Electroporation), calcium phosphate, DEAE-dextran/polybrene, viral based techniques and lipid based techniques.

[0642] The genetic material and/or nucleic acid to be delivered can be, by way of non-limiting example, nucleic acids that repair damaged or missing genes, nucleic acids for research applications, nucleic acids that kill a dysfunctional cell such as a cancer cell, antisense oligonucleotides to reduce or inhibit expression of a gene product, genetic material that increases expression of another gene, nucleotides and nucleotide analogs, peptide nucleic acids (PNAs), tRNAs, rRNAs, catalytic RNAs, RNA:DNA hybrid molecules, and combinations thereof.

[0643] The genetic material may comprise a gene expressing a protein. Example proteins include, but are not limited to, receptors (e.g., GPCRs, sphingolipid receptors, neurotransmitter receptors, sensory receptors, growth factor receptors, hormone receptors, chemokine receptors, cytokine receptors, immunological receptors, and compliment receptors, FC receptors), channels (e.g., potassium channels, sodium channels, calcium channels), pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potassium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidase), enzymes and kinases (e.g., protein kinases, ATPases, GTPases, phosphatases, proteases), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN), chemotactic/adhesion proteins (e.g., ICAM1, selectins, CD34, VCAM-1, LFA-1, VLA-1), and chimeric/fusion proteins (e.g., proteins in which a normally soluble protein is attached to a transmembrane region of another protein).

[0644] A minicell that is used to deliver therapeutic agents may comprise and display a binding moiety. By way of non-limiting example, binding moieties used for particular purposes may be a binding moiety directed to a compound or moiety displayed by a specific cell type or cells found predominantly in one type of tissue, which may be used, among other things, to target minicells and their contents to specific cell types or tissues. A preferred binding moiety is an antibody or antibody derivative. Other binding moieties include, but are not limited to, receptors, enzymes, ligands, binding peptides, fusion proteins, small molecules conjugated to transmembrane proteins, ligands conjugated to transmembrane proteins, viral fusion proteins, and fusion/chimeric proteins.

[0645] A minicell containing genetic material may be to a target cell by methods including, but not limited to, receptor mediated endocytosis, cell fusion, or phagocytosis (Aderem et al., Mechanism of Phagocytosis in Macrophages, Annu. Rev. Immunol. 17:593-623, 1999). The minicell gene delivery system is used to deliver genetic material in culture for research applications as well as to cells in vivo as part of gene therapy or other therapeutic applications.


[0647] Another non-limiting example of the minicell gene delivery and transfection system using invasin involves the expression of invasin following a targeting event. In this example, a minicell expresses a targeting protein that is capable of bringing the minicell in contact with a specific target cell. Upon contact with the target cell, the minicell will be induced to transcribe and translate invasin. The induction is accomplished via signaling events or with a transcription factor dimerization event. The minicells can be engineered to contain targeting proteins that induce protein expression only upon contact with a specific target cell. By way of non-limiting example, the invasin is expressed only at the target cell where it induces endocytosis, thus preventing the minicell from entering any cell but the target cell.

[0648] Proteins can be induced and expressed post contact with target cells include but are not limited to antibodies and antibody derivatives, receptors, enzymes, ligands, binding peptides, fusion proteins, small molecules conjugated to transmembrane proteins, ligands conjugated to transmembrane proteins, viral fusion proteins, antibiotics, apoptotic proteins, hormones, toxins, poisons, and fusion/chimeric proteins.

[0649] Another non-limiting example of gene delivery or transfection using the minicell involves the use of the type III secretion apparatus of bacteria. The type III secretion apparatus is expressed in the minicell and used to transfer genetic material to a target cell.

[0650] Another non-limiting example of gene delivery and transfection using minicells involves minicells that have been engineered to contain anionic lipids or cationic lipids (Axel et al., “Toxicity, Uptake Kinetics and Efficacy of New Transfection Reagents: Increase of Oligonucleotide Uptake,” Jour. of Vase. Res. 040:1-14, 2000). Many types of lipids have been shown to induce or enhance transfection and gene delivery in
a variety of cell types. Minicells containing such lipids could be used to transfer genetic material to specific cell types. Minicells can also be engineered to express targeting proteins that would allow the minicell to associate tightly with a target cell, which will facilitate the lipid interactions and gene transfer.

Another non-limiting example of gene delivery or transfection using minicells involves the use of ligands to induce receptor mediated endocytosis. By way of non-limiting example, the ligand is expressed on the surface of the minicell, or is attached to the surface of the minicell. A minicell containing genetic material is then able to associate with a target cell expressing the target receptor for the ligand. The receptor/ligand interaction will result in the endocytosis of the minicell into the target cell where the minicell would release and deliver the genetic material.

Another non-limiting example of gene delivery or transfection using minicells involves the use of fusion proteins, such as but not limited to viral capsid proteins. In this example the fusion protein would be expressed or attached to the outside of the minicell. The fusion protein would then induce fusion of the target cell with the minicell on contact. The contact could be initiated via random non-targeting events or via the use of specific targeting proteins. In both cases the end result would be the fusion of the minicell with a target cell and the delivery of the genetic material.

Another non-limiting example of research application of minicells involves their use for metabolic studies. The minicell can be generated to express metabolic pathways and the kinetics and function of the pathway can be studied.

Another non-limiting example of research application for minicells involves uses in cell free production of functional proteins (Jerntus et al., Recent advances in producing and selecting functional proteins by using cell-free translation, Current Opinion in Biotechnology 9:534-548, 1999). Minicells can be prepared as a reagent used to prepare compositions for in vitro translation. As is described in detail elsewhere herein, the composition of minicells can be manipulated so as to be enriched for particular proteins or nucleic acids, including those involved in protein translation and folding and/or modification of the proteins so produced into functional forms, i.e., forms having the activity of the corresponding protein as it is isolated from natural sources.

Non-limiting examples of such proteins and nucleic acids are ribosomal RNAs, ribosomal proteins, tRNAs, and the like.

Minicells could also be used in manual, semi-automated, automated and/or robotic assays for the in vitro determinations of the compounds of interest including, by way of non-limiting example, ligands, proteins, small molecules, bioactive lipids, drugs, heavy metals, and the like in environmental samples (e.g., air, water, soil), blood, urine or tissue of humans or samples from non-human organisms (e.g., plants, animals, protists) for the purpose of quantifying one or more compounds in a sample. A non-limiting example of this type of research application is the expression on the surfaces of the minicells of a receptor such as the receptor that binds a toxin produced by Baccillus anthracis. The protein, protective antigen (PA), is a 82.7 kDa protein that binds one of the secreted anthrax toxins, lethal factor (LF) (see Price, B. et al., Infection and Immunity 69: 4509-4515, 2001). Minicells expressing the PA protein could be used to detect LF in an environmental sample or in human blood, urine or tissue for the purposes of determining the presence of anthrax. As a non-limiting example, a competitive binding assay or an antibody-based assay could be used to indicate binding of LF in the environmental or tissue sample. Another non-limiting example is the use of PA-expressing minicells in a lateral flow diagnostic where interaction between the minicells and the LF-containing sample is indicated by the presence of a colored reaction product on a test strip.

Minicell-Based Delivery of Biologically Active Agents

General Considerations

The minicells of the invention are capable of encapsulating and/or loading into a membrane a variety of substances, including but not limited to biologically active agents, including but not limited to diagnostic and therapeutic agents. Biologically active agents include, but are not limited to, nucleic acids, e.g., DNA, RNA, gene therapy constructs, ribozymes, antisense and other synthetic oligonucleotides including those with chemical modifications; peptide nucleic acids (PNAs); proteins; synthetic oligopeptides; peptidomimetics; small molecules; radioisotopes; antibiotics; antibodies and antibody derivatives; and combinations and/or prodrugs of any of the preceding.

The surface of a minicell may be chemically altered in order to have certain properties that are desirable for their use as drug delivery agents. By way of non-limiting example, minicells may be chemically conjugated to polyethylene glyco-
col (PEG), which provides for "stealth" minicells that are not taken as well and/or as quickly by the reticuloendothelial system (RES). Other compounds that may be attached to minicells include without limitation polysaccharides, polynucleotides, lipopolysaccharides, lipoproteins, glycosylated proteins, synthetic chemical compounds, and/or combinations of any of the preceding.

[0670] A minicell that is used to deliver therapeutic agents may comprise and display a binding moiety. By way of non-limiting example, binding moieties used for particular purposes may be a binding moiety directed to a compound or moiety displayed by a specific cell type or cells found predominantly in one type of tissue, which may be used, among other things, to target minicells and their contents to specific cell types or tissues. A preferred binding moiety is an antibody or antibody derivative, which are described in detail elsewhere herein. Other binding moieties include, but are not limited to, receptors, enzymes, ligands, binding peptides, fusion proteins, small molecules conjugated to transmembrane proteins, ligands conjugated to transmembrane proteins, viral fusion proteins, and fusion/chimeric proteins.

[0671] XIII.B. Cellular Uptake

[0672] In addition to binding moieties, proteins and other compounds that induce or enhance the uptake or fusion of the minicell with the target gene can be displayed on the surface of a minicell for applications involving the delivery of therapeutic agents, gene therapy, and/or transfection or other research applications. See, generally, Adhesion Protein Protocols, Vol. 96, Dejuna, E. and Cortada, M., eds., Humana Press, 1999.

[0673] XIII.B.1. Cellular Uptake Sequences from Eukaryotic Cells

[0674] Eukaryotic adhesion receptors, which mediate intercellular adhesion, can be used as agents or targets for cellular uptake. There are at least three distinct classes of adhesive molecules that leukocytes employ during their adhesive interactions (a) integrins, including but not limited to LEC-CAMS/Selectins (ELAM-1, LAM-1/LReB/ITQL, and GMP140/PADGEM); (b) those belonging to the immunoglobulin superfamily including but not limited to CD2/LFA-2, CD3/TCR, CD4, CD8, CD28, CD44, CD54 (ICAM-1), ICAM-2, CD58 (LFA-3), VCAM-1, B7; and (c) Class I and II Major Histocompatibility Antigens (MHC).


[0676] By way of non-limiting example, a minicell display an adhesion receptor, or a fusion protein that has a transmembrane domain linked to a functional portion of an adhesion receptor. Such minicells will bind to cells displaying the ligand for the adhesion receptor.

[0677] XIII.B.2. Cellular Uptake Sequences from Prokaryotes


[0679] By way of non-limiting example, a minicell may express a protein such as invasin to induce receptor mediated endocytosis (Pepe et al., Versinia enterococitica invasin: A primary role in the initiation of infection, Proc. Natl. Acad. Sci. U.S.A. 90:6473-6477, 1993; Alrutz et al., Involvement of focal adhesion kinase in invasin-mediated uptake, Proc. Natl. Acad. Sci. U.S.A. 95:13658-13663, 1998). Invasin interacts with the Beta2 Integrin protein and causes it to dimerize. Upon dimerization the Beta2 Integrin signals for an endocytic event. Thus a minicell expressing the invasin protein will be taken up by cells expressing Beta2 Integrin via endocytosis.

[0680] As another non-limiting example, the pneumococcal adhesion protein CbpA interacts with the human polymyoglobulin receptor (hplgR) as either a part of the outer surface of a bacterial cell or as a free molecule Zhang et al. (Cell 102:827-837, 2000). The regions of CbpAhplgR interaction were mapped using a series of large peptide fragments derived from CbpA. CbpA (Swiss-Prot Accession No. O30874) contains a choline binding domain containing residues 454-663 and two N-terminal repetitive regions called R1 and R2 that are contained in residues 97-203 and 259-365, respectively. Polypeptides containing R1 and R2 interact with hplgR, whereas polypeptides containing other sequences from CbpA do not bind to hplgR. The R1 and/or R2 sequences of the CbpA polypeptide, and/or essentially iden-
tical, substantially identical, or homologous amino acid sequences, are used to facilitate the uptake of minicells by cells.

[0681] Another non-limiting example of gene delivery or transfection using the minicell involves the use of the type III secretion apparatus of bacteria. The type III secretion apparatus is expressed in the minicell and used to transfer genetic material to a target cell.

[0682] Other non-limiting examples of a minicell gene delivery and transfection targeting moiety are ETA (detoxified exotoxin A) protein delivery element described in U.S. Pat. No. 6,086,900 to Draper; Interalin and related proteins from Listeria species (Galan, Alternative Strategies for Becoming an Insider: Lessons from the Bacterial World, Cell 103:363-366, 2000); Intimin from pathogenic E. coli strains (Frankel et al., Intimin and the host cell—is it bound to end in Tir(s)? Trends in Microbiology 9:214-218); and SpeB, streptococcal pyrogenic exotoxin B (Stockbauer et al., A natural variant of the cysteine proteinase virulence factor of group A Streptococcus with an arginine-glycine-aspartic acid (RGD) motif preferentially binds human integrins αβ3 and α7β3, Proc. Natl. Acad. Sci. U.S.A. 96:242-247, 1999).

[0683] XIII.3. Cellular Uptake Sequences from Viruses

[0684] Cellular uptake sequences derived from viruses include, but are not limited to, the VP22 protein delivery element derived from herpes simplex virus-1 and vectors containing sequences encoding the VP22 protein delivery element are commercially available from Invitrogen (Carlsbad, Calif.; see also U.S. Pat. No. 6,017,735 to O hare et al.); and the Tat protein delivery element derived from the amino acid sequence of the Tat protein of human immunodeficiency virus (HIV). See U.S. Pat. Nos. 5,804,604; 5,747,641; and 5,674,980.

[0685] XIII.3.4. Lipids

[0686] Another non-limiting example of gene delivery and transfection using minicells involves minicells that have been engineered to contain anionic lipids or cationic lipids (Axel et al., Toxicity, Uptake Kinetics and Efficacy of New Transfection Reagents: Increase of Oligonucleotide Uptake, Jour. of Vasc. Res. 040:1-14, 2000). Many types of lipids have been shown to induce or enhance transfection and gene delivery in a variety of cell types. Minicells containing such lipids could be used to transfer genetic material to specific cell types. Minicells can also be engineered to express targeting proteins that would allow the minicell to associate tightly with a target cell, which will facilitate the lipid interactions and gene transfer.

[0687] Another non-limiting example of gene delivery or transfection using minicells involves the use of ligands to induce receptor mediated endocytosis. By way of non-limiting example, the ligand is expressed on the surface of the minicell, or is attached to the surface of the minicell. A minicell containing genetic material is then able to associate with a target cell expressing the receptor ligand for the ligand. The receptor/ligand interaction will result in the endocytosis of the minicell into the target cell where the minicell would release and deliver the genetic material.

[0688] Another non-limiting example of gene delivery or transfection using minicells involves the use of fusion proteins, such as but not limited to viral capsid proteins. In this example the fusion protein would be expressed or attached to the outside of the minicell. The fusion protein would then induce fusion of a target cell with the minicell upon contact. The contact could be initiated via random non-targeting events or via the use of specific targeting proteins. In both cases the end result would be the fusion of the minicell with a target cell and the delivery of the genetic material.

[0689] XIII.C. Post-Targeting Expression of Cellular Uptake Sequences

[0690] Another non-limiting example of the minicell gene delivery and transfection system using invasin involves the expression of invasin following a targeting event. In this example, a minicell expresses a targeting protein that is capable of bringing the minicell in contact with a specific target cell. Upon contact with the target cell, the minicell will be induced to transcribe and translate invasin. The induction is accomplished via signaling events or with a transcription factor dimerization event. The minicells can be engineered to contain targeting proteins that induce protein expression only upon contact with a specific target cell. By way of non-limiting example, the invasin is expressed only at the target cell where it induces endocytosis, thus preventing the minicell from entering any cell but the target cell.

[0691] Proteins can be induced and expressed post contact with target cells include but are not limited to antibodies and antibody derivatives, receptors, enzymes, ligands, binding peptides, fusion proteins, small molecules conjugated to transmembrane proteins, ligands conjugated to transmembrane proteins, viral fusion proteins, antibiotics, apoptotic proteins, hormones, toxins, poisons, and fusion/chimeric proteins.

[0692] XIII.D. Intracellular Targeting and Organellar Delivery

[0693] After delivery to and entry into a targeted cell, a minicell may be designed so as to be degraded, thereby releasing the therapeutic agent into the cytoplasm of the cell. The minicell and/or therapeutic agent may include one or more organellar delivery elements, which targets a protein into or out of a specific organelle or organelles. For example, the ricin A chain can be included in a fusion protein to mediate its delivery from the endosome into the cytosol. Additionally or alternatively, delivery elements for other organelles or subcellular spaces such as the nucleus, nucleolus, mitochondria, the Golgi apparatus, the endoplasmic reticulum (ER), the cytoplasm, etc. are included Mammalian expression constructs that incorporate organellar delivery elements are commercially available from Invitrogen (Carlsbad, CA: pShooter™ vectors). An H/KDEL (i.e., His/Lys-Asp-Glu-Leu sequence) may be incorporated into a fusion protein of the invention, preferably at the carboxy-terminus, in order to direct a fusion protein to the ER (see Andres et al., J. Biol. Chem. 266:14277-14278, 1991; and Pelham, Trends Bio. Sci. 15:483-486, 1990).

[0694] Another type of organellar delivery element is one which directs the fusion protein to the cell membrane and which may include a membrane-anchoring element. Depending on the nature of the anchoring element, it can be cleaved on the internal or external leaflet of the membrane, thereby delivering the fusion protein to the intracellular or extracellular compartment, respectively. For example, it has been demonstrated that mammalian proteins can be linked to i) myristic acid by an amide-linkage to an N-terminal glycine residue, to ii) a fatty acid or diacylglycerol through an amide- or thioether-linkage of an N-terminal cysteine, respectively, or covalently to iii) a phosphotidylinositol (PI) molecule through a C-terminal amino acid of a protein (for review, see Low, Biochem. J. 244:1-13, 1987). In the latter case, the PI molecule is linked to the C-terminus of the protein through an
interacting glycan structure, and the PI then embeds itself into the phospholipid bilayer; hence the term “GPI” anchor. Specific examples of proteins known to have GPI anchors and their C-terminal amino acid sequences have been reported (see Table I and FIG. 4 in Low, Biochimica et Biophysica Acta, 988:427-454, 1989; and Table 3 in Ferguson, Ann. Rev. Biochem., 57:285-320, 1988). Incorporation of GPI anchors and other membrane-targeting elements into the amino- or carboxy-terminus of a fusion protein can direct the chimeric molecule to the cell surface.

[XIII.E. Minicell-Based Gene Therapy]

[0695] The delivery of nucleic acids to treat diseases or disorders is known as gene therapy (Kay et al., Gene Therapy, Proc. Natl. Acad. Sci. USA 94:12744-12746, 1997). It has been proposed to use gene therapy to treat genetic disorders as well as pathogenic diseases. For reviews, see Desnick et al., Gene Therapy for Genetic Diseases, Acta Paediatr. Jpn. 40:191-203, 1998; and Bunnell et al., Gene Therapy for Infectious Diseases, Clinical Microbiology Reviews 11:42-56, 1998).

[0697] Gene delivery systems use vectors that contain or are attached to therapeutic nucleic acids. These vectors facilitate the uptake of the nucleic acid into the cell and may additionally help direct the nucleic acid to a preferred site of action, e.g., the nucleus or cytoplasm (Wu et al., “Delivery Systems for Gene Therapy,” Biotherapy 3:87-95, 1991). Different gene delivery vectors vary with regards to various properties, and different properties are desirable depending on the intended use of such vectors. However, certain properties (for example, safety, ease of preparation, etc.) are generally desirable in most circumstances.

[0698] The minicells of the invention may be used as delivery agents for any therapeutic or diagnostic agent, including without limitation gene therapy constructs. Minicells that are used as delivery agents for gene therapeutics constructs may, but need not be, targeted to specific cells, tissues, organs or systems of an organism, of a pathogen thereof, using binding moieties as described in detail elsewhere herein.


[0702] Minicells are used to deliver DNA-based gene therapy to targeted cells and tissues. Double minicell transformants are used not only to target a particular cell/tissue type (e.g. HIV-infected T-cells) but are also engineered to fuse with and enter targeted cells and deliver a protein-based toxin (e.g., antibiotic, or pro-apoptotic gene like Bax), an antisense expression construct (e.g., antisense to a transcription factor), or antisense oligonucleotides (e.g., antisense to an anti-apoptotic gene such as Bcl-2). The doubly-transformed minicells express not only these cell death promoters, but also only target particular cells/tissues, thus minimizing toxicity and lack of specificity of gene therapy vectors. By “doubly-transformed” it is meant that the minicell comprises 2 expression elements, one eubacterial, the other eukaryotic. Alternatively, shuttle vectors, which comprise eubacterial and eukaryotic expression elements in one vector, may be used.

[0703] Minicell-based gene therapy is used to deliver expression plasmids that could correct protein expression deficiencies or abnormalities. As a non-limiting example, minicell inhalants are targeted to pulmonary alveolar cells and are used to deliver chloride transporters that are deficient or otherwise material in cystic fibrosis. Protein hormone deli-
ciencies (e.g., dwarfism) are corrected by minicell expression systems (e.g., growth hormone expression in pituitary cells). Duchenne’s muscular dystrophy is characterized by a mutation in the dystrophin gene; this condition is corrected by minicell-based gene therapy. Angiogenesis treatment for heart patients is made effective by FGF or VEGF-producing minicells targeted to the heart. In this case, plasmid-driven over-expression of these grown factors is preferred.

XIV. Therapeutic Uses of Minicells

[0704] In addition to minicell-based gene therapy, minicells can be used in a variety of therapeutic modalities. Non-limiting examples of these modalities include the following applications.

XIV.A. Diseases and Disorders

[0706] Diseases and disorders to which the invention can be applied include, by way of non-limiting example, the following:

[0708] Diseases and disorders that involve the respiratory system, such as cystic fibrosis, lung cancer and tumors, asthma, pathogenic infections, allergy-related diseases and disorders, such as asthma; allergic bronchopulmonary aspergillosis; hypersensitivity pneumonia, eosinophilic pneumonia; emphysema; bronchitis; allergic bronchitis; bronchiectasis; cystic fibrosis; hypersensitivity pneumonitis; occupational asthma; sarcoid; reactive airway disease syndrome, interstitial lung disease, hyper-eosinophilic syndrome, parasitic lung disease and lung cancer; asthma, adult respiratory distress syndrome, and the like;

[0709] Diseases and disorders of the digestive system, such as those of the gastrointestinal tract, including cancers, tumors, pathogenic infections, colitis, ulcerative colitis, diverticulitis, Crohn’s disease, gastroenteritis, inflammatory bowel disease, bowel surgery ulceration of the duodenum, a mucosal villous disease including but not limited to coeliac disease, post-infective villous atrophy and short gut syndromes, pancreatitis, disorders relating to gastrointestinal hormones, Crohn’s disease, and the like;

[0710] Diseases and disorders of the skeletal system, such as spinal muscular atrophy, rheumatoid arthritis, osteoarthritis, osteoporosis, multiple myeloma-related bone disorder, cortical-ostearthritic-spinal degeneration, and the like;

[0711] Autoimmune diseases, such as Rheumatoid arthritis (RA), multiple sclerosis (MS), Sjogren’s syndrome, sarcoidosis, insulin dependent diabetes mellitus (IDDM), autoimmune thyroiditis, reactive arthritis, ankylosing spondylitis, scleroderma, polyomysitis, dermatomyositis, psoriasis, vasculitis, Wegener’s granulomatosis, Crohn’s disease and ulcerative colitis amyotrophic lateral sclerosis, multiple sclerosis, autoimmune gastritis, systemic lupus erythematosus, autoimmune hemolytic anemia, autoimmune neutropenia, systemic lupus erythematosus, graft vs. host disease, bone marrow engraftment, some cases of Type I diabetes, and the like;

[0712] Neurological diseases and disorders, such as depression, bipolar disorder, schizophrenia, Alzheimer’s disease, Parkinson’s disease, familial tremors, Gilles de la Tourette syndrome, eating disorders, Lewy-body dementia, chronic pain and the like;

[0713] Pathological diseases and resultant disorders such as bacterial infections such as infection by Escherichia, Shigella, Salmonella; sepsis, septic shock, and bacteremia; infections by a virus such as HIV, adenovirus, smallpox virus, hepatovirus, and the like; and AIDS-related encephalitis, HIV-related encephalitis, chronic active hepatitis, and the like;

[0714] Proliferative disease and disorders, such as acute lymphoblastic leukemia, acute myelogenous leukemia, chronic myelogenous leukemia, metastatic melanoma, Kaposi’s sarcoma, multiple myeloma, breast cancer, anal cancer, vulvar cancer, and the like; and

[0715] Various diseases, disorders and traumas including, but not limited to, apoptosis mediated diseases, inflammation, cerebral ischemia, myocardial ischemia, aging, sarcoidosis, granulomatous colitis, scleroderma, degenerative diseases, neoplastic diseases, alopecia, neurological damage due to stroke, diffuse cerebral cortical atrophy, Pick disease, mesolimbicocortical dementia, thalamic degeneration, Huntington chorea, cortical-basal ganglionic degeneration, cerebrocerebellar degeneration, familial dementia with spastic paraparesis, polyglucosan body disease, Shy-Drager syndrome, olivopontocerebellar atrophy, progressive supranuclear palsy, dystonia musculorum deformans, Hallervorden-Spatz disease, Meige syndrome, acanthocytic chorea, Friedreich ataxia, Holmes familial cortical cerebellar atrophy, Gerstmann-Straussler-Scheinker disease, progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, hereditary muscular atrophy, spastic paraplegia, glomerulonephritis, chronic thyroiditis, Grave’s disease, thymocytopenia, myasthenia gravis, psoriasis, peroneal muscular atrophy, hypertrophic interstitial polynuropathy, hereditary ataxic polyneuropathies, optic neuropathy, and ophthalmoplegia.

[0716] A variety of diseases and disorders caused or exacerbated by pathogens may be treated using the invention. For a comprehensive description of pathogens and associated diseases and disorders, see Zinsser Microbiology, 20th Ed., Joklik, ed., Appleton-Century-Crofts, Norwalk, Conn., 1992, and references cited therein.

[0717] Minicells could also be used for replacement therapy (via gene therapy) in a variety of conditions known to be caused by protein or proteins that are either absent (e.g. Duchene’s Muscular Dystrophy), reduced in level (Dwarfism) or abberant (Sickle-cell anemia).


XIV.B. Removal of Toxins and Pathogens by Selective Absorption

[0720] When introduced into the bloodstream of an animal, receptor-displaying minicells bind and absorb toxic compounds, thereby removing such compounds from the general circulation. A therapeutic benefit ensues as the bound toxic compound cannot access the cells upon which it would otherwise exert its toxic effect.

[0721] Minicells expressing receptors for toxic substances are introduced IV in order to remove those toxins from the blood. One non-limiting example is in the treatment of sepsis. In one embodiment, a fusion protein is formed from the transmembrane domain of the FGF receptor or toxR and a known soluble receptor for LPS (lipopolysaccharide), such as the LBP (lipopolysaccharide binding protein) or the extracellular domain of CD14 receptor protein, both of which bind the LPS bacterial endotoxin. These minicells inactivate LPS by initially binding to it and preventing LPS binding to naturally occurring CD14 receptors on heart cells and other cells
involved in the endotoxic shock response. Eventually, the minicell-LPS complex is cleared from the blood by macrophages and other components of the immune system.

[0722] In another embodiment, minicells expressing receptors for toxic drugs (e.g., morphine) are used to treat drug overdoses. In other embodiments, minicells of the invention are used to express receptors to venoms (e.g., snake venom) or poisons (e.g., muscarinic receptor expression for the treatment of muscarine poisoning). In other embodiments, minicells of the invention expressing EDGRs are used to clear the blood of toxins and other undesirable compounds.

[0723] As another non-limiting example, minicells that bind pathogens are used to treat disease. Minicells, and pathogens bound thereto, may be ingested by human neutrophils and thus serve as an adjuvant to therapeutic processes mediated by neutrophils (Fox et al., Fate of the DNA in plasmid-containing Escherichia coli minicells ingested by human neutrophils, Blood 69:1394-400, 1987). In a related modality, minicells are used to bind compounds required for the growth of a pathogen.

[0724] XIV.C. Antiviral Therapy

[0725] In one modality, minicells of the invention are used as “sponges” for the selective absorption of any viral particle in the body. Without being limited to the following examples, minicells expression receptors or antibodies selectively directed against viruses such as HIV, Hepatitis B and smallpox are used.

[0726] For the treatment of viremia, viruses are cleared from the blood by absorption during dialysis or by IV injection of minicells expressing targets for viral receptors. As the minicells interact with blood-borne virus particles, there is an initial reduction of host cell infection by virtue of the minicell-viral complexes that are formed. Since viral particles attach to and/or enter the minicell, they are not active because of the lack of machinery needed for viral replication in the minicells. The virus infected minicells are then cleared from the system by macrophages and processed by the immune system.

[0727] Certain retroviruses that infect particular host cells express viral proteins on the surfaces of the infected cells. HIV infection of T-cells is one non-limiting example of this. The viral protein, GP120, is expressed on the surfaces of infected T-cells (Turner et al., Structural Biology of HIV, J. Mol. Biol. 285:1-52, 1999). Minicells expressing CD4 act as anti-GP120 minicells not only to target virus particles in an infected patient, but also to identify infected T-cells. It may be desirable to also express co-receptors such as CCR5, CXCR4 or ARD (Dragic, An overview of the determinants of CCR5 and CXCR4 co-receptor function, J. Gen. Virol. 82:1807-1814, 2001). The minicells are then used to kill the infected T-cells, or to inhibit viral replication and/or virion assembly.

[0728] In another non-limiting example of anti-pathogen therapy, minicells can be used to express bacterial surface antigens on their surfaces that facilitate cellular uptake of the minicell by intracellular pathogens such as Mycobacterium tuberculosis (the causative agent of tuberculosis), Rickettsiae, or viruses. In this “smart sponge” approach, selective absorption is accompanied by internalization of the pathogen by minicells. Destruction of the pathogen follows as a result of a combination of intraminicell digestion of pathogens and/or by the eventual processing of the virus-containing minicell by the cellular immune system of the patient.

[0729] XIV.D. Antibacterial and Antiparasitic Applications

[0730] Minicells may be used to kill pathogenic bacteria, protozoans, yeast and other fungi, parasitic worms, viruses and other pathogens by mechanisms that either do or do not rely on selective absorption. Antibiotics can be delivered to pathogenic organisms after being targeted by the proteins or small molecules on the surfaces of the minicells that promote binding of the minicells to the surfaces of the pathogen. Fusion or injection of minicell contents into the pathogenic cell can result in the death or disablement of the pathogen and thus lower the effective dose of an antibiotic or gene therapeutic agent. Delivery of antibiotics tethered to or encapsulated by the minicells will reduce the effective dose of an antibiotic and will reduce its elimination by the renal system. In the case of delivering encapsulated molecules (e.g., antibiotics), purified/isolated minicells expressing membrane-bound proteins for targeting can be incubated with the molecules in vitro prior to administration. This would be particularly applicable to the use of protoplast minicells or poroplast minicells that have their outer membrane and cell wall or outer membrane only removed, respectively, thus facilitating the diffusion of the small molecule into the intact minicell.

[0731] Without being limited by the following example, minicells can be used as antibacterial agents by expressing on the surfaces of the minicells antigens, receptors, antibodies, or other targeting elements that will target the minicell to the pathogenic organism and facilitate the entry of plasmids, proteins, small molecules in order to gain access to or entry into the organism. Antibiotics may be encapsulated by minicells post isolation from the parent strain so that the antibiotic will not be effective against the minicell-producing bacteria itself. Since minicells are not able to reproduce, the antibiotic will not be lethal to the minicell delivery vehicle, but only to the targeted pathogen. In another non-limiting example, lysogenic factors e.g., complement may be expressed on the surfaces of the minicells or encapsulated by the same as to promote lysis of the pathogen.

[0732] Minicells can also be engineered to express toxic proteins or other elements upon binding to the pathogen. Induction of minicell protein expression can be an event that is coincident with targeting or triggered by minicell binding to the target pathogen, thus making minicells toxic only when contact is made with the pathogenic organism. Minicells can be engineered to express fusion/chimeric proteins that are tethered to the membrane by transmembrane domains that have signaling moieties on the cytoplasmic surfaces of the minicells, such as kinases or transcription factors. In one non-limiting example, a minicell fusion membrane-bound protein could be expressed containing an extracellular domain with a receptor, scFv, or other targeting protein that binds to the pathogen. The second segment of the chimera could be a transmembrane domain of a protein such as the EGF receptor or ToxR (that would tether the fusion protein to the membrane). Importantly, the cytoplasmic domain of the fusion protein could be a kinase that phosphorylates a bacterial transcription factor present in the minicell or could be fused to a transcription factor that would be expressed on the cytoplasmic surface of the minicell. The expression plasmid that was previously introduced into the minicells would then be activated by promoters utilizing the activated bacterial transcription factor pre-existing in the minicells or that which may be introduced by the minicell. Without being limited to the following example, the binding event could be signaled by a fusion protein containing elements of a receptor (e.g., EGF).
or by an adhesion protein (e.g., an integrin) that require oligomerization. In the example of the use of integrins, bacterial or other transcription factors that also require dimerization could be cloned as fusion proteins such that the binding event would be signaled by a dimerization of two or more identical recombinant chimeric proteins that have association-dependent transcription factors tagged to the C-terminus of the fusion protein. The minicells would only be toxic when contact is made with the pathogen.

The proposed mechanism of induction coincident with targeting is not limited to the antiparasitic uses of minicells but can be used in other therapeutic situations where minicells are used to express proteins of therapeutic benefit when directed against eucaryotic cells of the organism (e.g., kill cancer cells with protein toxins expressed only after binding of the minicell to the cancer cell).

Transfer of DNA-containing plasmids or other expression elements, antisense genes, etc. may be used to express toxic proteins in the target cells or otherwise inhibit transcript ion and/or translation in the pathogenic organism or would otherwise be toxic to the cell. Without being limited by the following example, minicells can be used to transfer plasmids expressing growth repressors, DNAs, or other proteins or peptides (e.g., pro-apoptotic) that would be toxic to the pathogen.

Fusion proteins expressed in minicells are used for cancer therapy. In a non-limiting example, phage display antibody libraries are used to clone single chain antibodies against tumor-associated (tumor-specific) antigens, such as Much-1 or EGFVIII. Fusion proteins expressing these antibodies, and further comprising a single-pass transmembrane domain of an integral membrane protein, are used to “present” the antibody to the surface of the minicells. Injected minicells coated with anti-tumor antibodies target the tumor and deliver pro-apoptotic genes or other toxic substances to the tumor. The minicells are engulfed by the tumor cells by processes such as receptor-mediated endocytosis (by, e.g., macrophages). By way of non-limiting example, toxin-invasive could be expressed on the surfaces of the minicells to promote endocytosis through the interaction between invasion and beta2-integrins on the surfaces of the target cells.

Fusion proteins possessing viral fusion-promoting proteins facilitate entry of the minicell to the tumor cell for gene therapy or for delivery of chemotherapeutics. In these and similar applications, the minicell may contain separate eukaryotic and bacterial expression elements, or the expression elements may be combined into a single “shuttle vector.”

Minicells as diagnostic tools can be used either in vitro or in vivo. In the in vitro context, the minicells are used in an ELISA format in a lateral flow diagnostic platform to detect the presence and level of a desired analyte. A sample (tissue, cell or bodily fluid sample) is taken and then tested in vitro. One advantage of the minicell system in detecting substances in tissue, cells or in body fluids is that the minicells can be used in vitro assays where the minicell is labeled with either a radioactive or fluorescent compound to aid in its detection in an ELISA format or lateral flow platform. Thus, the use of secondary antibody detection systems is obviated.

As an in vivo diagnostic, minicells can be radiolabeled. One method of labeling is to incubate minicells for a short time (about 8 hr) with a I-125 tracer (e.g., Tn99M) that is useful for detecting tumor metastases. The Tn99M accumulates in cells and loads into minicells after isolation or into the parent bacteria during growth phase. As Tn99M is oxidized by either the parent E. coli strain or by the minicells after isolation, the Tn99M is retained by the cell. Iodine-labeled proteins may also be used (Krown et al., TNF-alpha receptor expression in rat cardiac myocytes: TNF-alpha inhibition of L-type Ca2+ transients, FEBS Letters 376:24-30, 1995).

One non-limiting example of in vivo detection of cancer making use of radiolabeled minicells is the use of the minicells to express chimeric membrane-bound single-chain fusion protein that acts as a membrane bound receptor for a particular ligand. The fusion protein then acts as a membrane bound receptor for a particular ligand or molecule in the sample. Conventional cell cloning techniques (e.g., PCR) are used to identify genes for binding proteins, or phage display is used to identify a gene for a single-stranded variable antibody gene expressing binding protein for a particular ligand. The protein product is preferably a soluble protein. By constructing a plasmid containing this gene plus the transmembrane domain of a known single-pass membrane protein such as that of the EGF receptor, a fusion protein may be expressed on the surfaces of the minicells as an integral membrane protein with an extracelular domain that is preferably capable of binding ligand.

In another type of fusion protein, the transmembrane domain of the EGF receptor is fused to a known soluble receptor for a particular ligand, such as the LBP (lipopolysaccharide binding protein) or the extracellular domain of CD14 receptor protein, both of which bind the bacterial endotoxin, LPS (lipopolysaccharide). The LBP/EGF or CD14/EGF fusion protein is used to measure LPS in the serum of patients suspected of sepsis.

The minicell system is used to express receptors such as those of the EDG (endothelial cell differentiation gene) family (e.g., EDG 1-9) that recognize sphingolipids such as sphingosine-1-phosphate (SIP), sphingosylphosphorylcholine (SPC) and the lysosphospholipid, lysosphosphatidic acid (LPA). Since these proteins are 7-pass integral membrane proteins, no additional transmembrane domains of another protein are needed, and the receptor protein is thus not a fusion protein.

Truncated or mutant forms of a protein of interest are useful in a diagnostic assay. For example, a protein that is an ligand-binding enzyme can be altered so as to bind its substrate of interest but can no longer convert substrate into product. One example of this application of minicell technology is the expression of a truncated or mutant lactase dehydrogenase which is able to bind lactic acid, but is not able to convert lactic acid to pyruvate. Similarly, hexokinase derivatives are used in minicells for glucose monitoring.

Minicells as diagnostic tools can be used either in vitro or in vivo. In the in vitro context, the minicells are used in an ELISA format or in a lateral flow diagnostic platform to detect the presence and level of a desired analyte. A sample (tissue, cell or bodily fluid sample) is taken and then tested in vitro. One advantage of the minicell system in detecting substances in tissue, cells or in body fluids is that the minicells can be used in vitro assays where the minicell is labeled with either a radioactive or fluorescent compound to aid in its detection in a lateral flow platform. Thus, the use of secondary antibody detection systems is obviated.

As an in vivo diagnostic, minicells can be radiolabeled. One method of labeling is to incubate minicells for a short time (about 8 hr) with a I-125 tracer (e.g., Tn99M) that is useful for detecting tumor metastases. The Tn99M accumulates in cells and loads into minicells after isolation or into the parent bacteria during growth phase. As Tn99M is oxidized by either the parent E. coli strain or by the minicells after isolation, the Tn99M is retained by the cell. Iodine-labeled proteins may also be used (Krown et al., TNF-alpha receptor expression in rat cardiac myocytes: TNF-alpha inhibition of L-type Ca2+ transients, FEBS Letters 376:24-30, 1995).

One non-limiting example of in vivo detection of cancer making use of radiolabeled minicells is the use of the minicells to express chimeric membrane-bound single-chain...
antibodies against tumor-specific antigens (TSA) expressed on malignant melanoma or other transformed cells. Such TSAs include, but are not limited to, the breast cancer associated MUC1 antigen and variant forms of the EGFR (EGF-FvIII). By way of non-limiting example, minicells expressing antibodies to melanoma cells can be injected (IV) into a patient and then subjected to CAI scan of the lymphatic drainage in order to determine if a metastasis has occurred. This diagnostic technique obviates the need for lymph node dissection.

Another example of an in vivo diagnostic is to use the minicell system to express antibodies against oxidized low-density lipoproteins (LDL). Oxidized LDLs are associated with atherogenic plaques. Radiolabeled minicells (prepared as above) are injected IV into a person prior to nuclear imaging for image enhancement. MRI image contrast enhancement is performed by preparing minicells complexed (loaded) with contrast enhancers such as paramagnetic relaxivity agents and magnetic susceptibility agents.

In diagnostic as well as other applications, minicells preferentially detect a diagnostic marker, i.e., a marker associated with a disease or disorder. A diagnostic marker is statistically more likely to occur in individuals suffering from a disease than in those who are not diseased. Preferably, a diagnostic marker directly causes or is produced during a disease; however, the association may be no more than a correlation.

XVI. Drug Discovery (Screening) with Minicells

XVI.A. Assays

Minicells can be used in assays for screening pharmacological agents. By way of non-limiting example, the minicell system provides an environment for the expression of GPCRs and studies of their ligand binding kinetics. Such GPCR’s include any member the Endothelial Differentiation Gene (EDG) receptor family. GPCRs may participate in neoplastic cell proliferation, angiogenesis and cell death. Small molecules that either activate or inhibit the action of these GPCRs can be used in therapeutic interaction.

Assays are performed to determine protein expression and protein function. For example, the production of the protein can be followed using protein 35S-Met labeling. This is performed by providing the cell only methionine that is labeled with 35S. The cells are treated with IPTG to induce protein expression, and the 35S-Met is incorporated into the protein. The cells are then lysed, and the resulting lysates are electrophoresed on an SDS gel and exposed to autoradiography film.

Another technique for assessing protein expression involves the use of western blots. Antibodies directed to various expressed proteins of interest have been generated and are commercially available. Techniques for generating antibodies to proteins or polypeptides derived therefrom are known in the art (see, e.g., Cooper et al., Section III of Chapter 11 in: Short Protocols in Molecular Biology, 2nd Ed., Ausubel et al., eds., John Wiley and Sons, New York, 1992, pages 11-22 to 11-46). Standard western blot protocols, which may be used to show protein expression from the expression vectors in minicells and other expression systems, are known in the art. (see, e.g., Winston et al., Unit 10.7 of Chapter 10 in: Short Protocols in Molecular Biology, 2nd Ed., Ausubel et al., eds., John Wiley and Sons, New York, 1992, pages 10-32 to 10-35).

The amount of functional protein produced from a minicell expression system is determined through the use of binding studies. Ligands for the proteins of interest are used to show specific binding in the minicell system. Radiolabeled ligand is incubated with cells expressing the protein, in this case, a receptor for TNF-alpha. The cells are then centrifuged and the radioactivity counted in a scintillation counter. The amount of ligand that is bound reflects the amount of functional protein that is present in the sample.

By way of non-limiting example, the minicell system can be made to express EDGRs for the purpose of screening combinatorial chemistry libraries for molecules that enhance EDG activity. EDG activity is assayed in the minicell environment in several ways. One way is to crystalize minicells expressing an EDG protein (or any membrane-bound protein of choice) and then measure changes in the crystal structure to detect novel ligands. Circular dichroism (CD), x-ray diffraction, electron spin resonance (ESR) or other biophysical approaches are used to probe the structure of proteins in the minicell context. Additionally or alternately, minicells are produced that express not only the EDGR, but also express G-proteins (i.e., double transfectants). An assay system involving GTP binding and hydrolysis is used to identify and assess which small molecules in the combinatorial chemistry library serve as activating ligands for EDG. The minicell expression system is used in in vitro binding assays and in high throughput drug screenings. The expression of mutant, truncated, or chimeric isoforms of proteins are used for functional analyses in order to discover inactive or overactive proteins for potential use in diagnostics or therapeutics.

XVI.B. High-Throughput Screening (HTS)

HTS typically uses automated assays to search through large numbers of compounds for a desired activity. Typically HTS assays are used to find new drugs by screening for chemicals that act on a particular enzyme or molecule. For example, if a chemical inactivates an enzyme it might prove to be effective in preventing a process in a cell that causes a disease. High throughput methods enable researchers to try out thousands of different chemicals against each target very quickly using robotic handling systems and automated analysis of results.

As used herein, “high throughput screening” or “HTS” refers to the rapid in vitro screening of large numbers of compounds (libraries); generally tens to hundreds of thousands of compounds, using robotic screening assays. Ultra high-throughput Screening (uHTS) generally refers to the high-throughput screening accelerated to greater than 100,000 tests per day.

To achieve high-throughput screening, it is best to house samples on a multicontainer carrier or platform. A multicontainer carrier facilitates measuring reactions of a plurality of candidate compounds simultaneously. Multi-well microplates can be used as the carrier. Such multi-well microplates, and methods for their use in numerous assays, are both known in the art and commercially available.

Screening assays may include controls for purposes of calibration and confirmation of proper manipulation of the components of the assay. Blank wells that contain all of the reactants but no member of the chemical library are usually included. As another example, a known inhibitor (or activator) of an enzyme for which modulators are sought, can be incubated with one sample of the assay, and the resulting decrease (or increase) in the enzyme activity determined according to the methods herein. It will be appreciated that modulators can also be combined with the enzyme activators or inhibitors to find modulators which inhibit the enzyme
activation or repression that is otherwise caused by the presence of the known enzyme modulator. Similarly, when ligands to a sphingolipid target are sought, known ligands of the target can be present in control/calibration assay wells.

The minicells of the invention are readily adaptable for use in high-throughput screening assays for screening candidate compounds to identify those which have a desired activity, e.g., inhibiting an enzyme that catalyzes a reaction that produces an undesirable compound, inhibiting function of a receptor independent of ligand interference, or blocking the binding of a ligand to a receptor therefor. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as therapeutic agents.

The methods of screening of the invention comprise using screening assays to identify, from a library of diverse molecules, one or more compounds having a desired activity. A "screening assay" is a selective assay designed to identify, isolate, and/or determine the structure of, compounds within a collection that have a preselected activity. By "identifying" it is meant that a compound having a desirable activity is isolated, its chemical structure is determined (including without limitation determining the nucleotide and amino acid sequences of nucleic acids and polypeptides, respectively) the structure of and, additionally or alternatively, purifying compounds having the screened activity). Biochemical and biological assays are designed to test for activity in a broad range of systems ranging from protein-protein interactions, enzyme catalysis, small molecule-protein binding, agonists and antagonists, to cellular functions. Such assays include automated, semi-automated assays and HTS (high throughput screening) assays.

In HTS methods, many discrete compounds are preferably tested in parallel by robotic, automatic or semi-automatic methods so that large numbers of test compounds are screened for a desired activity simultaneously or nearly simultaneously. It is possible to assay and screen up to about 6,000 to 20,000, and even up to about 100,000 to 1,000,000 different compounds a day using the integrated systems of the invention.

Typically in HTS, target molecules are contained in each well of a multi-well microplate; in the case of enzymes, reactants are also present in the wells. Currently, the most widely established techniques utilize 96-well microtiter plates. In this format, 96 independent tests are performed simultaneously on a single 8 cm x 12 cm plastic plate that contains 96 reaction wells. One or more blank wells contains all of the reactants except the candidate compound. Each of the non-standard wells contain at least one candidate compound.

These wells typically require assay volumes that range from 50 to 500 ul. In addition to the plates, many instruments, materials, pipettors, robotics, plate washers and plate readers are commercially available to fit the 96-well format to a wide range of homogeneous and heterogeneous assays. Microtiter plates with more wells, such as 384-well microtiter plates, are also used, as are emerging methods such as the nanowell method described by Schullek et al. (Anal Biochem., 30 246, 20-29, 1997).

In one modality, screening comprises contacting a sphingolipid target with a diverse library of member compounds, some of which are ligands of the target, under conditions where complexes between the target and ligands can form, and identifying which members of the libraries are present in such complexes. In another non-limiting modality, screening comprises contacting a target enzyme with a diverse library of member compounds, some of which are inhibitors (or activators) of the target, under conditions where a product or a reactant of the reaction catalyzed by the enzyme produce a detectable signal. In the latter modality, inhibitors of target enzyme decrease the signal from a detectable product or increase a signal from a detectable reactant (or vice-versa for activators).

Minicells of the invention expressing and/or displaying a protein are used for screening assays designed to identify agents that modulate the activity of the target protein. Such assays include competitive inhibition binding assays for high throughput assays. Competitive inhibition assays include but are not limited to assays that screen agents against a specific target protein to identify agents that inhibit, interfere, block, or compete with protein-ligand interactions, protein-protein interactions, enzymatic activity, or function of a specific protein. Examples of competitive inhibition include but are not limited to the development of neutral inhibitors of the serine protease factor Xa that were discovered using a high throughput screening assay using a compound library (Carr et al., Neutral inhibitors of the serine protease factor Xa, Bioorg Med Chem Lett 11, 2001), the design and characterization of potent inhibitors for the human oxytocin receptor (Seyer et al, Design, synthesis and pharmacological characterization of a potent radio iodinated and photoactivatable peptide oxytocin antagonist, J Med Chem. 44:3022-30, 2001), and the identification of non-peptide somatostatin antagonists of the sst(3) protein (Thurieau et al, Identification of potent non-peptide somatostatin agonists with sst(3) selectivity, J Med Chem. 44:2990-3000, 2001).

High throughput competitive inhibition assays are designed to identify agents that inhibit a specific target protein. Such assays include but are not limited to ones that measure enzymatic activity, protein-ligand interactions, protein-protein interactions and other functions of proteins. Minicells that express and/or display a specific protein could be used in all types of competitive inhibition assays.

One non-limiting example of high throughput competitive inhibition screening using minicells for the purpose of this patent involves the competitive inhibition of known ligands. The ligand is attached to but not limited to a fluorophore, fluorescent protein, tags such as 6xHis tag or FLAG tag, chromophores, radiolabeled proteins and molecules, binding moieties such as avidin and strepavidin, voltage sensitive dyes and proteins, bioluminescent proteins and molecules, or fluorescent peptides. The target protein, which binds the tagged ligand, is expressed and stably displayed by the minicell. When the ligand is added to the minicell solution the ligand binds to the target protein. Following a wash the interaction is detected via the fluorophore, fluorescent protein, tag, or fluorescent peptide. The ligand-bound minicells could either be centrifuged (taking advantage of the sedimentation properties of the minicell particle) or immunoprecipitated with an antibody against an antigen expressed on the minicell membrane or the minicells can be adsorbed/fixed to a substrate such as a standard 96 well plate. The competitive inhibition assay is carried out by adding agents to the minicell mix either before, together or after the ligand is added. Thus if the agent is a competitive inhibitor of the ligand to the target protein the ligand will be washed away from the minicell because it is not associated with the target protein. The agent prevents binding and thus eliminated the detection signal from the minicell.
Minicells of this invention are used in "functional screening HTS assays". Functional screening assays are defined as assays that provide information about the function of a specific target protein. Functional assays screen agents against specific target proteins to identify agents that either act as an antagonist or as an agonist against the protein. Functional assays require that the target protein be in an environment that allows it to carry out its natural function. Such functions include but are not limited to G-proteins coupling with a GPCR, enzymatic activity such as phosphorolysis or proteolysis, protein-protein interaction, and transport of molecules and ions.

Functional assays screen agents against proteins which are capable of natural function. Target proteins used in functional studies must carry out a functional assay that is measurable. Examples of protein functions that are measurable include but are not limited to the use of Fluorescent Resonance Energy Transfer (FRET) to measure the G-protein coupling to a GPCR (Ruiz-Velasco et al., Functional expression and FRET analysis of green fluorescent proteins fused to G-protein subunits in rat sympathetic neurons, J Physiol. 537:679-692, 2001; Janetopoulos et al., Receptor-mediated activation of heterotrimeric G-proteins in living cells, Science 291:2408-2411, 2001); Bioluminescence Resonance Energy Transfer (BRET) to assay for functional ligand induced G-protein coupling to a target GPCR (Menard, L. Bioluminescence Resonance Energy Transfer (BRET): A powerful platform to study G protein coupled receptors (GPCR) activity in intact cells, Assay Development, Nov. 28-30, 2001), the use of fluorescent substrates to measure the enzymatic activity of proteases (Grant, Designing biochemical assays for protein cleavage using fluorogenic substrates, Assay Development, Nov. 28-30, 2001); and the determination of ion channel function via the use of voltage sensitive dies (Andrews et al., Correlated measurements of free and total intracellular calcium concentration in central nervous system neurons, Microsc Res Tech. 46:370-379, 1999).

One non-limiting example of high throughput functional screening assay using minicells for the purpose of this patent involves the functional coupling of GPCRs to their respective G-protein. Upon ligand binding, voltage polarization, ion binding, light interaction and other stimulatory events activate GPCRs and cause them to couple to their respective G-protein. In a minicell, both the GPCR and its respective G-proteins can be simultaneously expressed. Upon activation of the GPCR the coupling event will occur in the minicell. Thus by detecting this coupling in the minicell, one could screen for agents that bind GPCRs to identify antagonists and agonists. The agonists are identified by screening for agents that activate the GPCR in the absence of the natural activator.

The detection of GPCR activation and coupling in a minicell is accomplished by using systems that generate a signal upon coupling. One non-limiting example involves the use of BRET or FRET. These systems require that two fluorescent or bioluminescent molecules or proteins be brought into close contact. Thus by attaching one of these molecules or proteins to the GPCR and one to the G-protein, they will be brought together upon coupling and a signal will be generated. This signal can be detected using specific detection equipment and the coupling event can be monitored. Thus the function of the GPCR can be assayed and used in functional assays in minicells.

Another non-limiting functional assay for GPCRs and other proteins in minicells involves the use of transcription factors. Many bacterial transcription factors and eukaryotic transcription factors require dimerization for activation. By attaching one subunit of a transcription factor to a GPCR and the other subunit to a G-protein, the subunits will dimerize upon coupling of the GPCR to the G-protein because they will be brought into close contact. The dimerized transcription factor will then be activated and will act on its target episomal DNA. In the minicell system the episomal DNA target will be a plasmid that encodes for proteins that provide a signal for detection. Such proteins include but are not limited to luciferase; green fluorescent protein (GFP), and derivatives thereof such as YFP, BFP, etc.; alcohol dehydrogenase, and other proteins that can be assayed for expression. The activation of the GPCR will result in coupling and activation of the transcription factor. The transcription factor will then induce transcription and translation of specific detector proteins. Thus the activation of the GPCR will be monitored via the expression of the detector protein.

In another modality, the transcription factor can inhibit expression in the minicell system and thus allowing for the screening of constitutively active GPCRs and proteins. For example if the GPCR were constitutively active then the transcription factor to use would be one that inhibits transcription and translation. Thus agents could be screened against the constitutively active GPCR to identify agents that caused the constitutively active GPCR to uncouple. The uncoupling will result in the inactivation of the transcription factor. The inhibition caused by the transcription factor will be removed and transcription and translation will occur. Thus a detectable protein will be made and a signal will be received.

The transcription dimerization assay can be used for any protein function that involves a protein-protein interaction, protein-ligand interaction and protein-drug interaction. Thus any assay involving such interactions can be carried out in the minicell.

Another non-limiting functional screening assay involves the use of enzymatic function to screen for functionality. In this modality the receptor or other protein performs a specific enzymatic function. This function is then carried out in the minicell and monitored using biochemical and other techniques. For example if the target protein was a protease then fluorescent peptides with the cleavage site of the protease could be used to monitor the activity of the protease. If the protease was functioning then the peptide would be cleaved and the fluorescent would change. Thus agents can be screened against the protease in the minicell and the fluoroscents can be monitored using specific detection systems. In another non-limiting example, a membrane-bound enzyme such as sphingomyelinase could be expressed in minicells and the minicell particles adsorbed to a standard substrate such as a 96 well plate. The enzymatic activity could be assessed by a standard in vitro assay involving the release of product (phosphocholine) (e.g., Amplex® kit A-12220 sold by Molecular Probes). Sphingomyelinase inhibitors could be screened by measuring the reduction of phosphocholine production in the well when presented with substrate (sphingomyelin) in a coupled fluorescence assay.

The detection of GPCR activation and coupling in a minicell is accomplished by using systems that generate a signal upon coupling. One non-limiting example involves the use of BRET or FRET. These systems require that two fluorescent or bioluminescent molecules or proteins be brought into close contact. Thus by attaching one of these molecules or proteins to the GPCR and one to the G-protein, they will be brought together upon coupling and a signal will be generated. This signal can be detected using specific detection equipment and the coupling event can be monitored. Thus the function of the GPCR can be assayed and used in functional assays in minicells.

Another non-limiting functional assay for GPCRs and other proteins in minicells involves the use of transcription factors. Many bacterial transcription factors and eukaryotic transcription factors require dimerization for activation. By attaching one subunit of a transcription factor to a GPCR and the other subunit to a G-protein, the subunits will dimerize upon coupling of the GPCR to the G-protein because they will be brought into close contact. The dimerized transcription factor will then be activated and will act on its target episomal DNA. In the minicell system the episomal DNA target will be a plasmid that encodes for proteins that provide a signal for detection. Such proteins include but are not limited to luciferase; green fluorescent protein (GFP), and derivatives thereof such as YFP, BFP, etc.; alcohol dehydrogenase, and other proteins that can be assayed for expression. The activation of the GPCR will result in coupling and activation of the transcription factor. The transcription factor will then induce transcription and translation of specific detector proteins. Thus the activation of the GPCR will be monitored via the expression of the detector protein.

In another modality, the transcription factor can inhibit expression in the minicell system and thus allowing for the screening of constitutively active GPCRs and proteins. For example if the GPCR were constitutively active then the transcription factor to use would be one that inhibits transcription and translation. Thus agents could be screened against the constitutively active GPCR to identify agents that caused the constitutively active GPCR to uncouple. The uncoupling will result in the inactivation of the transcription factor. The inhibition caused by the transcription factor will be removed and transcription and translation will occur. Thus a detectable protein will be made and a signal will be received.

The transcription dimerization assay can be used for any protein function that involves a protein-protein interaction, protein-ligand interaction and protein-drug interaction. Thus any assay involving such interactions can be carried out in the minicell.

Another non-limiting functional screening assay involves the use of enzymatic function to screen for functionality. In this modality the receptor or other protein performs a specific enzymatic function. This function is then carried out in the minicell and monitored using biochemical and other techniques. For example if the target protein was a protease then fluorescent peptides with the cleavage site of the protease could be used to monitor the activity of the protease. If the protease was functioning then the peptide would be cleaved and the fluorescent would change. Thus agents can be screened against the protease in the minicell and the fluoroscents can be monitored using specific detection systems. In another non-limiting example, a membrane-bound enzyme such as sphingomyelinase could be expressed in minicells and the minicell particles adsorbed to a standard substrate such as a 96 well plate. The enzymatic activity could be assessed by a standard in vitro assay involving the release of product (phosphocholine) (e.g., Amplex® kit A-12220 sold by Molecular Probes). Sphingomyelinase inhibitors could be screened by measuring the reduction of phosphocholine production in the well when presented with substrate (sphingomyelin) in a coupled fluorescence assay.
Another non-limiting example of minicells used for functional assays involves the screening of agonists/antagonists for ion channels. In this example the calcium channel, SCAamper, is encoded on a poycistrionic episomal plasmid, which also encodes for a luminaceous soluble protein, aequorin. In this assay, the minicell will contain aequorin protein in its cytoplasm and SCAamper proteins expressed on the minicell membrane. Thus upon activation of SCAamper by its ligand, SPC, or by an analog thereof, calcium will flow into the minicell and will be bound by the aequorin which will luminesce. Thus a detection signal for the functional activation of the calcium channel is obtained.

Minicells can also be employed for expression of target proteins and the preparation of membrane preparations for use in screening assays. Such proteins include but are not limited to receptors (e.g., GPCRs, sphingolipid receptors, neurotransmitter receptors, sensory receptors, growth factor receptors, hormone receptors, chemokine receptors, cytokine receptors, immunological receptors, and compliment receptors, FC receptors), channels (e.g., potassium channels, sodium channels, calcium channels), pores (e.g., nuclear pore proteins, water channels), ion and other pumps (e.g., calcium pumps, proton pumps), exchangers (e.g., sodium/potassium exchangers, sodium/hydrogen exchangers, potassium/hydrogen exchangers), electron transport proteins (e.g., cytochrome oxidase), enzymes and kinases (e.g., kinases, ATPases, GTPases, phosphatases, proteases), structural/linker proteins (e.g., Caveolins, clathrin), adapter proteins (e.g., TRAD, TRAP, FAN), chemotactic/adhesion proteins (e.g., ICAM1, selectins, CD34, VCAM-1, LFA-1, VLA-1), and chimeric/fusion proteins (e.g., proteins in which a normally soluble protein is attached to a transmembrane region of another protein). In such assays the membrane preparations are used to screen for agents that are either antagonists or agonists. These assays use various formats including but not limited to competitive inhibition.

The format for the screening of minicells includes but is not limited to the use of test tubes, 6 well plates, 12 well plates, 24 well plates, 96 well plates, 384 well plates, 1536 well plates, and other microtiter plate well plates. In these systems the minicells can be immobilized, attached, bound, or fused with the above test tubes or plates. The minicells can also be free in solution for use in tubes and plates. The detection systems for the minicell assay include but are not limited to fluorescent plate readers, scintillation counters, spectrophotometers, Viewlux CCD Imager, Luminex, AIPHAQuest, BLAscoring, FLIPR and F-MAT. Minicell assays can be carried out with but not limited to techniques such as manual handling, liquid handlers, robotic automated systems and other formats.

XVI.C. Chemical Libraries

Developments in combinatorial chemistry allow the rapid and economical synthesis of hundreds to thousands of discrete compounds. These compounds are typically arrayed in moderate-sized libraries of small organic molecules designed for efficient screening. Combinatorial methods, can be used to generate unbiased libraries suitable for the identification of novel inhibitors. In addition, smaller, less diverse libraries can be generated that are descendened from a single parent compound with a previously determined biological activity. In either case, the lack of efficient screening systems to specifically target therapeutically relevant biological molecules produced by combinatorial chemistry such as inhibitors of important enzymes hampers the optimal use of these resources.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical “building blocks,” such as reagents. For example, a linear combinatorial chemical library, such as a polypeptide library, is formed by combining a set of chemical building blocks (amino acids) in a large number of combinations, and potentially in every possible way, for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

A “library” may comprise from 2 to 50,000,000 diverse member compounds. Preferably, a library comprises at least 48 diverse compounds, preferably 96 or more diverse compounds, more preferably 384 or more diverse compounds, more preferably 10,000 or more diverse compounds, preferably more than 100,000 diverse members and most preferably more than 1,000,000 diverse member compounds. By “diversity” it is meant that greater than 50% of the compounds in a library have chemical structures that are not identical to any other member of the library. Preferably, greater than 75% of the compounds in a library have chemical structures that are not identical to any other member of the collection, more preferably greater than 90% and most preferably greater than about 99%.


Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka, Int. J. Pept. Prot. Res., 37:487-493 (1991) and Houghton et al., Nature, 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to, peptide (PCT Publication No. WO91/179735); encoded peptides (PCT Publication WO 93/20242); random bio-oligomers (PCT Publication No. WO 92/00091); benzodiazepines (U.S. Pat. No. 5,288,514); diversomers, such as hydantoin, benzodiazepines and dipeptides (Hobbs et al., Proc. Nat. Acad. Sci. USA, 90:6909-6913 (1993); vinylogous polypeptides


[0787] XVI.D. Measuring Enzymatic and Binding Reactions During Screening Assays

[0788] Techniques for measuring the progression of enzymatic and binding reactions in multicontainer carriers are known in the art and include, but are not limited to, the following.

[0789] Spectrophotometric and fluorometric assays are well known in the art. Examples of such assays include the use of colorimetric assays for the detection of peroxides, as disclosed in Example 1(b) and Gordon, A. J. and Ford, R. A., The Chemist’s Companion: A Handbook Of Practical Data, Techniques, And References, John Wiley and Sons, N.Y., 1972, Page 437.

[0790] Fluorescence spectrometry may be used to monitor the generation of reaction products. Fluorescence methodology is generally more sensitive than the absorption methodology. The use of fluorescent probes is well known to those skilled in the art. For reviews, see Bushford et al., Spectrophotometry and Spectrofluorometry: A Practical Approach, pp. 91-114, IRL Press Ltd. (1987) and Bell, Spectroscopy In Biochemistry, Vol. I, pp. 155-194, CRC Press (1981).

[0791] In spectrofluorometric methods, enzymes are exposed to substrates that change their intrinsic fluorescence when processed by the target enzyme. Typically, the substrate is nonfluorescent and converted to a fluorophore through one or more reactions. As a non-limiting example, SMase activity can be detected using the Amplex® Red reagent (Molecular Probes, Eugene, Oreg.). In order to measure sphingomyelinase activity using Amplex Red, the following reactions occur. First, SMase hydrolyzes sphingomyelin to yield ceramide and phosphocholine. Second, alkaline phosphatase hydrolyzes phosphorylcholine to yield choline. Third, choline is oxidized by choline oxidase to betaine. Finally, H₂O₂ in the presence of horseradish peroxidase, reacts with Amplex Red to produce the fluorescent product, Resorufin, and the signal therefrom is detected using spectrofluorometry.

[0792] Fluorescence polarization (FP) is based on a decrease in the speed of molecular rotation of a fluorophore that occurs upon binding to a larger molecule, such as a receptor protein, allowing for polarized fluorescent emission by the bound ligand. FP is empirically determined by measuring the vertical and horizontal components of fluorophore emission following excitation with plane polarized light. Polarized emission is increased when the molecular rotation of a fluorophore is reduced. A fluorophore produces a larger polarized signal when it is bound to a larger molecule (i.e. a receptor), slowing molecular rotation of the fluorophore. The magnitude of the polarized signal relates quantitatively to the extent of fluorescent ligand binding. Accordingly, polarization of the “bound” signal depends on maintenance of high affinity binding.

[0793] FP is a homogeneous technology and reactions are very rapid, taking seconds to minutes to reach equilibrium. The reagents are stable, and large batches may be prepared, resulting in high reproducibility. Because of these properties, FP has proven to be highly automatable, often performed with a single incubation with a single, premixed, tracer-receptor reagent. For a review, see Owicki et al., Application of Fluorescence Polarization Assays in High-Throughput Screening, Genetic Engineering News, 17:27, 1997.

[0794] FP is particularly desirable since its readout is independent of the emission intensity (Checovich, W. J., et al., Nature 375:254-256, 1995; Dandliker, W. B., et al., Methods in Enzymology 743:28, 1981) and is thus insensitive to the presence of colored compounds that quench fluorescence emission. FP and FRET (see below) are well-suited for identifying components that block interactions between receptors and their ligands. See, for example, Parker et al., Development of high throughput screening assays using fluorescence polarization: nuclear receptor-ligand-binding and kinase/phosphatase assays, J Biomol Screen 5:77-88, 2000.

[0795] Exemplary normal-and-polarized fluorescence readers include the POLARION fluorescence polarization system (Tecan A G, Hombrechtikon, Switzerland). General multiwell plate readers for other assays are available, such as the VERSAMAX reader and the SPECTRAMAX multiwell plate spectrophotometer (both from Molecular Devices).

[0796] Fluorescence resonance energy transfer (FRET) is another useful assay for detecting interaction and has been described previously. See, e.g., Heim et al., Curr. Biol. 6:178-182, 1996; Mitra et al., Gene 173:13-17 1996; and Selvin et al., Meth. Enzymol. 246:300-345, 1995. FRET detects the transfer of energy between two fluorescent substances in close proximity, having known excitation and emission wavelengths. As an example, a protein can be expressed as a fusion protein with green fluorescent protein (GFP). When two fluorescent proteins are in proximity, such as when a protein specifically interacts with a target molecule, the resonance energy can be transferred from one excited molecule to the other. As a result, the emission spectrum of the sample shifts, which can be measured by a fluorometer, such as a MAX multimultiwell fluorometer (Molecular Devices, Sunnyvale Calif.).

[0797] Scintillation proximity assay (SPA) is a particularly useful assay for detecting an interaction with the target molecule. SPA is widely used in the pharmaceutical industry and has been described (Hanselman et al., J. Lipid Res. 38:2365-2373 (1997); Kahl et al., Anal. Biochem. 243:282-283.
[0798] The target molecule can be bound to the scintillator plates by a variety of well known means. Scintillator plates are available that are derivatized to bind to fusion proteins such as GST, His6 or Flag fusion proteins. Where the target molecule is a protein complex or a multimer, one protein or subunit can be attached to the plate first, then the other components of the complex added later under binding conditions, resulting in a bound complex.

[0799] In a typical SPA assay, the gene products in the expression pool will have been radiolabeled and added to the wells, and allowed to interact with the solid phase, which is the immobilized target molecule and scintillator coating in the wells. The assay can be measured immediately or allowed to reach equilibrium. Either way, when a radiolabel becomes sufficiently close to the scintillator coating, it produces a signal detectable by a device such as a TopCount NXT microplate scintillation counter (Packard BioScience Co., Meriden Conn.). If a radiolabeled expression product binds to the target molecule, the radiolabel remains in proximity to the scintillator long enough to produce a detectable signal.

[0800] In contrast, the labeled proteins that do not bind to the target molecule, or bind only briefly, will not remain near the scintillator long enough to produce a signal above background. Any time spent near the scintillator caused by random Brownian motion will also not result in a significant amount of signal. Likewise, residual unincorporated radiolabel used during the expression step may be present, but will not generate significant signal because it will be in solution rather than interacting with the target molecule. These non-binding interactions will therefore cause a certain level of background signal that can be mathematically removed. If too many signals are obtained, salt or other modifiers can be added directly to the assay plates until the desired specificity is obtained (Nichols et al., Anal. Biochem. 257:112-119, 1998).

[0801] XVI. E. Screening for Novel Antibiotics

[0802] As bacteria and other pathogens acquire resistance to known antibiotics, there is an ongoing interest in identifying new antibiotics. See, e.g., Powell W A, Catranis C M, Maynard C A. Synthetic antimicrobial peptide design. Mol Plant Microbe Interact 1995 September-October; 8(5):792-4. Minicells can be used to assay, identify and purify novel antibiotics to eubacteria. By way of non-limiting example, a minicell that comprises a detectable compound can be contacted with a candidate antibiotic to see if the minicell is lysed by a candidate compound, which would release the detectable compound from the interior of the minicell into solution, this producing a signal that indicates that the candidate antibiotic is effective at lysing bacteria. In such assays, the detectable compound is such that it produces less or more of the same signal, or a different signal, inside the minicell as compared to in solution post-lysis. By way of non-limiting example, the minicell could comprise a fluorescent compounds that, when contacted with a second fluorescent compound in solution, produces FRET.

[0803] XVI.F. Reverse Screening

[0804] In one version of minicell display, the invention provides methods for screening libraries of minicells in which each minicell comprises an expression element that encodes a few, preferably one, membrane proteins in order to identify a membrane protein that interacts with a preselected compound. By way of non-limiting example, sequences encoding membrane proteins, fusion proteins, or cytoplasmic proteins are cloned into an expression vector, either by "shotgun" cloning or by directed cloning, e.g., by screening or selecting for cDNA clones, or by PCR amplification of DNA fragments, that encode a protein using one or more oligonucleotides encoding a highly conserved region of a protein family. For a non-limiting example of such techniques, see Krautwurst, D., et al. 1998. Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell 95:917-926. By way of non-limiting example, a minicell expressing a receptor binds a preselected ligand, which may be a drug. Various assays for receptor binding, enzymatic activity, and channeling events are known in the art and may include detectable compounds; in the case of binding assays, competition assays may also be used (Musimirembwa, C. M., et al. 2001. In vitro high throughput screening of compounds for favorable metabolic properties in drug discovery. Comb. Chem. High Throughput Screen. 4:245-263; Matthenkis, L. C., and A. Saychenko. 2001. Assay technologies for screening ion channel targets. Curr. Opin. Drug Discov. Devel. 4:124-134; Nammann, R., and P. A. Negulescu. 2001. High-throughput screening strategies for cardiac ion channels. Trends Cardiovasc. Med. 11:54-59; Le Poul, E., et al. 2002. Adaptation of aequorin functional assay to high throughput screening. J. Biomol. Screen. 7:57-65; and Graham, D. L., et al. 2001. Application of beta-galactosidase enzyme complementation technology as a high throughput screening format for antagonists of the epidermal growth factor receptor. J. Biomol. Screen. 6:401-411).

[0805] Once a minicell has been identified by an assay and isolated, DNA is prepared from the minicell. The cloned DNA present in the minicell encodes the receptor displayed by the minicell. Having been cloned, the receptor is used as a therapeutic target. For example, the receptor is produced via recombinant DNA expression and used in minicell-based or other assays to identify and characterize known and novel compounds that are ligands, antagonists and/or agonists of the cloned receptor. The ligands, antagonists and agonists may be used as lead compounds and/or drugs to treat diseases for which the drug is therapeutic are expected to be treated using the novel ligands, antagonists and agonists, or drugs and prodrugs developed therefrom.

[0806] Preparations of minicells that express and secrete secretes a soluble protein can be prepared in order to identify ligands, including but not limited to small molecules, that interact with the soluble protein. Soluble proteins include, but are not limited to, known secreted or proteolytically cleaved proteins and peptides, hormones and cytokines. In this format, minicells are placed in, or adhered to, the wells of a microtiter multiwell plate. A different compound or group of compounds is placed in each well, along with any reagents necessary to generate or squelch a signal corresponding to a change in the soluble protein produced by the minicell. Such changes include, by way of non-limiting example, conformational changes in the protein that may occur as a result of binding of a ligand or otherwise. A well that generates the appropriate signal contains a compound that causes a change in the soluble protein.

[0807] It is also possible to carry out procedures such as the one described in the immediately preceding paragraph “in
reverse.” In this format, a known ligand, which may be a drug, is used to identify soluble proteins that bind to the ligand/drug. Libraries of minicells wherein each minicell secretes a different soluble protein are prepared, and each type of minicell is placed into, or adhered to the wall of, a well of a microtiter plate, along with reagents for generating a signal when the ligand/drug binds to a soluble protein. Minicells that generate the appropriate signal comprise a cloned DNA that encodes a soluble protein that interacts with the known ligand/drug. Once cloned, the soluble protein is prepared and used as a therapeutic target in order to identify known or novel compounds that bind thereto. When the preselected ligand is a drug, diseases for which that drug is therapeutic are expected to be treated using the known and novel compounds so identified, or using drugs and prodrugs developed from such compounds.

Minicells expressing known membrane and soluble proteins can also be used to help characterize lead compounds and accelerate the generation of drugs therefrom. In particular, such studies may be used identify potentially detrimental interactions that might occur upon in vivo administration, e.g., ADME/Tox screening (Ekins, S., et al. 2002. In silico ADME/Tox: the state of the art. J. Mol. Graph. Model. 20:305-309; and Li, A., et al. 2002. Early ADME/Tox studies and in silico screening. Drug Discov. Today 7:25-27).

By way of non-limiting example, a human receptor that is known to be important for the normal functioning of a cell may be expressed in minicells, and various chemical derivatives of a lead compound can be tested to ensure that they do not bind to the receptor, as such binding would be expected to have adverse effects in vivo. As another example, an enzyme that degrades a drug, such as a cytochrome P450, is expressed in minicells and used to examine the susceptibility of a candidate drug to such degradation. The cytochrome P450 family of enzymes is primarily responsible for the metabolism of xenobiotics such as drugs, carcinogens and environmental chemicals, as well as several classes of endobiotics such as steroids and prostaglandins. Exemplary P450 cytochromes involved in drug degradation include, but are not limited to, CYP2D6 (cytochrome P4502D6, also known as debrisoquine hydroxylase), CYP1A1, CYP1A2, CYP2A6, CYP2B6, CYP2C9, CYP2C18, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP3A5.

In one aspect of the invention, minicells are used in methods of screening to identify agents that improve, enhance, or decrease the interaction of a protein with another compound. These methods include, by way of non-limiting example, modification of protein targets through directed or random mutagenic approaches to identify critical interactions between a wild-type protein target and a specific drug molecule. Information obtained from studies of mutant proteins is used to specifically produce or modify a therapeutic agent to interact more specifically and/or effectively with the wild-type protein target, thus increasing the therapeutic efficacy of the parental drug and/or decreasing non-specific, potentially deleterious interactions. See, for example, Lietha, D., et al. 2001. Crystal structures of NK1-heparin complexes reveal the basis for NK1 activity and enable engineering of potent agonists of the MET receptor. EMBO J. 20:5543-5555; and Chen, Y. Z., et al. Can an optimization/scoring procedure in ligand-protein docking be employed to probe drug-resistant mutations in proteins? J. Mol. Graph. Model. 19:560-570; Zhao, H. and F. H. Arnold. Combinatorial protein design: Strategies for screening protein libraries. Current Opinion in Structural Biology 7:480-485 (1997); and Cram PT, A, el Tayar N, Karlen A, Testa B. Molecular electrostatic potentials for characterizing drug-biosystem interactions. Methods Enzymol. 1991; 203:638-77. Martin Y C. Computer-assisted rational drug design. Methods Enzymol. 1991; 203:587-613.

By way of non-limiting example, information obtained using the methods of the invention may be in conjunction with x-ray crystallographic structural determinations to characterize receptor/ligand interactions (Müller, G. 2000. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. Curr. Med. Chem. 7:861-888). By way of non-limiting example, minicells may be used to display the family of molecular variants to characterize the specific mutagenic changes on the functional properties of the protein.

Studies of variant proteins can also be used to modify drugs to fit natural variants of proteins, especially those associated with pathogens. Pathogens such as viruses, including retroviruses such as HIV, may acquire mutations that change a site where a drug acts, thereby rendering the pathogen immune to the drug. Studies of variant proteins can be used to quickly produce derivatives of a drug that are active against a variant protein. See, for example, Vanghe J N, Smith P W, Sollis S L, Blick J T, Sahasrabudhe A, McKimm-Breschkin J L, Colman P M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. Structure 1998 Jun. 15; 6(6):735-46; and Baldwin E T, Bhat T N, Liu B, Pattabiraman N, Erickson J W. Structural basis of drug resistance for the V82A mutant of HIV-1 protease. 78: Nat Struct Biol 1995 March; 2(3):244-9.

XVI.H. Directed Evolution

The minicells and methods described herein can be used in directed evolution. Unlike natural variation, directed evolution generates new protein variants in vitro (see, e.g., Arnold, F. H. and A. A. Violett, Directed Evolution of Biocatalysts. Curr Op Chem Biol 1999. 3:54-59). Amino acid substitutions can be introduced into a protein of interest by mutating the gene encoding the protein. Mutations are introduced by, e.g., replicating DNA in mutator strains, by chemical mutagenesis or radiation-induced mutagenesis (Drake, J. W., The Molecular Basis of Mutation. Holden-Day, San Francisco, 1970). Other methods include error-prone PCR and “domain shuffling” (Moore, G. L. and C. D. Maranas, Modeling DNA Mutation and Recombination for Directed Evolution Experiments. J. Theor. Biol. 2000. 205:483-503). In the latter method, different regions of members of the same gene family are recombined so that the inherent variability of members of the family is used to produce novel “isofoms” of genes.

A group of variants is screened to select for those variants which have the desired activity. The activity of the initial variants that are so isolated may be inadequate for a given application, but the process can be repeated using these initial members to generate a second group of variants, or reiterated as many times as is necessary to produce one or more variants having the desired activity or characteristics.

XVI.I. Isolation and Characterization of Components of Signal Transduction Pathways

In one version of minicell display, the invention provides methods for screening libraries of minicells, in which each minicell comprises a preselected component of a signal transduction pathway, in order to identify soluble and
membrane proteins that interact with the preselected component. By way of non-limiting example, a plurality of minicells, each of which displays the same G-protein-coupled receptor (GPCR), is used to prepare a minicell library in which a different G-protein encoding sequence is present and expressed in each minicell. Minicells comprising a G-protein that interacts with the GPCR are identified, e.g., via transactivation assays described in Example 18. Once a minicell has been identified by an assay and isolated, DNA is prepared from the minicell. The cloned DNA present in the minicell encodes a G-protein that interacts with the GPCR of the displayed by the minicells of the library. Having been cloned, the G-protein is used as a therapeutic target that can be used in screening assays to identify novel lead compounds and drugs that interfere or alter the activity of the GPCR. In particular, when the GPCR of the minicell library is known to be a therapeutic target for a specific disease, it is expected that compounds that interfere or alter the activity of a G-protein that interacts with the GPCR will be lead to therapeutics for that specific disease.


[0820] XVII. Determining the Structures of Membrane Proteins

[0821] Three-dimensional (3D) structures of proteins may be used for drug discovery. However, GPCRs and other membrane proteins present challenging problems for 3D structure determination. Muller, Towards 3D structures of G-protein-coupled receptors: a multidisciplinary approach. (Review), Curr. Med Chem 2000 pp. 861-88; Levy et al., Two-dimensional crystallization on lipid layers. A successful approach for membrane proteins, J Struct Biol 1999 127, 44-52. Although the three-dimensional structures of hundreds of different folds of globular proteins have been determined, fewer than 20 different integral membrane protein structures have been determined. There are many reasons for this. Extracting membrane proteins from the membrane can easily disrupt their native structure, and membrane proteins are notoriously difficult to crystallize.

[0822] Some membrane proteins readily form two-dimensional crystals in membranes and can be used for structure determination using electron diffraction spectroscopy (ED) instead of x-ray crystallography. This is the technique that was used to determine the structure of bacteriorhodopsin (see below).

[0823] Nuclear magnetic resonance (NMR) is an alternative method for determining membrane protein structure, but most membrane proteins are too large for high-resolution NMR at the present state of the art. Furthermore, membrane proteins require special conditions for NMR, e.g. deuterated lipids must be used to avoid confusing the signal of the protein protons with the noise of membrane lipid protons.

<table>
<thead>
<tr>
<th>PROTEIN</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MONOTOPIC MEMBRANE PROTEINS</strong></td>
<td></td>
</tr>
<tr>
<td>Prostaglandin H2 synthase-1, Sheep, 3.5 Å</td>
<td>Picot et al. (1994)</td>
</tr>
<tr>
<td>Cyclooxygenase-2, Mus Musculus, 3.0 Å</td>
<td>Kamendulis et al. (1996)</td>
</tr>
<tr>
<td>Squalene-hopene cyclase, Alcyolocystis acidocaldarius, 2.0 Å</td>
<td>Wendt et al. (1999)</td>
</tr>
<tr>
<td><strong>TRANSMEMBRANE PROTEINS</strong></td>
<td></td>
</tr>
<tr>
<td>Bacteriophage Rhodopsin (Halobacterium salinarium)</td>
<td></td>
</tr>
<tr>
<td>Bacteriorhodopsin (BR)</td>
<td></td>
</tr>
<tr>
<td>2D xtal, EM, 3.5 Å</td>
<td>Grigorieff et al. (1996)</td>
</tr>
<tr>
<td>2D xtal, EM, 3.0 Å</td>
<td>Kiriura et al. (1997)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 2.5 Å</td>
<td>Pohmay-Peyrola et al. (1997)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 1.9 Å</td>
<td>Belhrali et al. (1999)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 2.1 Å K intermediate</td>
<td>Edman et al. (1999)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 2.3 Å</td>
<td>Luecke et al. (1998)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 1.55 Å</td>
<td>Luecke et al. (1999)</td>
</tr>
<tr>
<td>3D xtal, X-ray: D66N mutant (BR) 1.8 Å</td>
<td>Luecke et al. (1999)</td>
</tr>
<tr>
<td>3D xtal, X-ray: D66N mutant (M) 2.00 Å</td>
<td>Luecke et al. (1999)</td>
</tr>
<tr>
<td>3D xtal, X-ray, 2.9 Å</td>
<td>Essen et al. (1998)</td>
</tr>
<tr>
<td>Halorhodopsin (HR)</td>
<td></td>
</tr>
<tr>
<td>3D xtal, Xray, 1.8 Å</td>
<td>Kolbe et al. (2000)</td>
</tr>
<tr>
<td>G PROTEIN-COUPLED RECEPTORS</td>
<td></td>
</tr>
<tr>
<td>Rhodopsin, Bovine Rod Outer Segment, 2.8 Å</td>
<td>Palczewski et al. (2000)</td>
</tr>
<tr>
<td>Phototrophic Reaction Centers</td>
<td></td>
</tr>
<tr>
<td>Rhodopseudomonas viridis, 2.3 Å</td>
<td>Deisenhofer et al. (1985)</td>
</tr>
<tr>
<td>Rhodobacter sphaeroides, 3.0 Å</td>
<td>Yeates et al. (1987)</td>
</tr>
<tr>
<td>Rhodobacter sphaeroides, 3.1 Å</td>
<td>Chang et al. (1991)</td>
</tr>
<tr>
<td>Light Harvesting Complexes</td>
<td></td>
</tr>
<tr>
<td>Rhodopseudomonas acidophila, 2.5 Å</td>
<td>McDermott et al. (1995)</td>
</tr>
<tr>
<td>Rhodospirillum molischianum, 2.4 Å</td>
<td>Koepke et al. (1996)</td>
</tr>
<tr>
<td>Photosystems</td>
<td></td>
</tr>
<tr>
<td>Photosystem I, Synecococcus elongates 4.0 Å</td>
<td>Schubert et al. (1997)</td>
</tr>
<tr>
<td>Photosystem II, Synecococcus elongates 3.8 Å</td>
<td>Zunzi et al. (2001)</td>
</tr>
<tr>
<td>Beta-Barrel Membrane Proteins-Multimeric (Porins and Relatives)</td>
<td></td>
</tr>
<tr>
<td>Porin, Rhodobacter capsulatus, 1.8 Å</td>
<td>Weiss &amp; Schulz (1992)</td>
</tr>
<tr>
<td>Porin, Rhodopseudomonas blasticina 1.95 Å</td>
<td>Kreutzsch et al. (1994)</td>
</tr>
<tr>
<td>OmpF, E. coli, 2.4 Å</td>
<td>Cowan et al. (1992)</td>
</tr>
<tr>
<td>PhoE, E. coli, 3.0 Å</td>
<td>Cowan et al. (1992)</td>
</tr>
<tr>
<td>Maltoeprin, Salmonella typhimurium, 2.4 Å</td>
<td>Meyer et al. (1997)</td>
</tr>
<tr>
<td>Maltoeprin, E. coli, 3.1 Å</td>
<td>Schärer et al. (1995)</td>
</tr>
<tr>
<td>Beta-Barrel Membrane Proteins-Monomeric/Dimeric</td>
<td></td>
</tr>
<tr>
<td>TolC outer membrane protein, E. coli 2.1 Å Protein is a trimer, each contributing 4 strands to a single barrel.</td>
<td>Koronakis et al. (2000)</td>
</tr>
<tr>
<td>OmpA, E. coli, 2.5 Å</td>
<td>Paulus &amp; Schulz (1998)</td>
</tr>
<tr>
<td>OmpA, E. coli, By NMR, in DPC micelles</td>
<td>Rona et al. (2001)</td>
</tr>
<tr>
<td>OmpX, E. coli, 1.9 Å</td>
<td>Vogt &amp; Schulz (1990)</td>
</tr>
<tr>
<td>OMFLA (outer membrane phosphate lipase A) E. coli, 2.1 Å monomer (IQ95) and dimer (IQ96)</td>
<td>Stuiber et al. (1999)</td>
</tr>
<tr>
<td>FlaA, E. coli, 2.5 Å</td>
<td>Ferguson et al. (1998); Lambert et al., 1999</td>
</tr>
<tr>
<td>FlaA + ferriheme-iron, E. coli, 2.7 Å</td>
<td>Buchanan et al. (1999)</td>
</tr>
<tr>
<td>FepA, E. coli, 2.4 Å</td>
<td>Ferguson et al. (1999)</td>
</tr>
<tr>
<td>Glycerophorin A human</td>
<td>MacKenzie et al. (1997)</td>
</tr>
<tr>
<td>Non-constitutive Toxins, etc.</td>
<td></td>
</tr>
<tr>
<td>Alpha-hemolysin, Staphylococcus aureus, 1.9 Å</td>
<td>Song et al. (1996)</td>
</tr>
<tr>
<td>LukF, Staphylococcus aureus, 1.9 Å</td>
<td>Olson et al. (1999)</td>
</tr>
</tbody>
</table>
TABLE 5-continued

<table>
<thead>
<tr>
<th>PROTEIN</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ion Channels</strong></td>
<td></td>
</tr>
<tr>
<td>KcsA Potassium, H+ gated, Streptomyces lividans, 3.2 Å</td>
<td>Doyle et al. (1998)</td>
</tr>
<tr>
<td>Mcl, Mechanosensitive, Mycobacterium tuberculosis, 3.5 Å</td>
<td>Chang et al. (1998)</td>
</tr>
<tr>
<td><strong>Other Channels</strong></td>
<td></td>
</tr>
<tr>
<td>Electron crystallography in membrane plane. 3.8 Å</td>
<td>Ren et al. (2000)</td>
</tr>
<tr>
<td>AQP1 — In vitreous ice by electron microscopy, 3.7 Å</td>
<td>Fu et al. (2000)</td>
</tr>
<tr>
<td>GitP — Glycerol facilitator channel, E. coli, 2.2 Å</td>
<td></td>
</tr>
<tr>
<td>F-type ATPase</td>
<td></td>
</tr>
<tr>
<td><strong>Calcium ATPase, Sarcoplasmic reticulum, Rabbit, 2.6 Å</strong></td>
<td>Toyoshina et al. (2000)</td>
</tr>
<tr>
<td><strong>Respiratory Proteins</strong></td>
<td></td>
</tr>
<tr>
<td>Fumarate Reductase Complex, Escherichia coli, 3.3 Å</td>
<td>Iverson et al. (1999)</td>
</tr>
<tr>
<td>Fumarate Reductase Complex, Wolinella succinogenes, 2.2 Å</td>
<td>Lancaster et al. (1999)</td>
</tr>
<tr>
<td>ATP synthase (F,subg.), S. cerevisiae, 3.9 Å. X-ray structure is a C alpha model derived from composite of 1BMF, 1AQI &amp; 1AQY</td>
<td>Stock et al. (1999)</td>
</tr>
<tr>
<td><strong>Cytochrome C Oxidases</strong></td>
<td></td>
</tr>
<tr>
<td>aa, bovine heart mitochondria, 2.8 Å</td>
<td>Tsukihara et al. (1996)</td>
</tr>
<tr>
<td>aa, Paracoccus denitrificans, 2.8 Å</td>
<td>Iwata et al. (1995)</td>
</tr>
<tr>
<td>b5, from T. thermophilus, 2.4 Å</td>
<td>Soulimane et al. (2000)</td>
</tr>
<tr>
<td>Cytochrome b5, Complexes</td>
<td></td>
</tr>
<tr>
<td>Bovine Heart Mitochondria (5 subunits), 2.9 Å</td>
<td>Xia et al. (1997)</td>
</tr>
<tr>
<td>Chicken Heart Mitochondria, 3.16 Å</td>
<td>Zhang et al. (1998)</td>
</tr>
<tr>
<td>Bovine Heart Mitochondria (11 subunits), 2.8-3.0 Å. S. cerevisiae, (yeast, 9 subunits), 2.3 Å</td>
<td>Iwata et al. (1998)</td>
</tr>
<tr>
<td></td>
<td>Haute et al. (2000)</td>
</tr>
</tbody>
</table>

Citations for Table 5:  
Haute et al. (1999). Structure of 2.3 Å resolution of cytochrome b5 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. Structure 8, 669-684.  
Koch et al. (1996). The crystal structure of the light-harvesting complex II (B800-850) from Rhodospirillum molischianum. Structure 4, 581-597.  
TABLE 5-continued

<table>
<thead>
<tr>
<th>PROTEIN</th>
<th>REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodimuth et al. (2000). Structure and mechanism of the ubiquitin binding of cytochrome c oxidase from Thermus thermophilus. EMBO J. 19, 1766-1776.</td>
<td></td>
</tr>
<tr>
<td>Tsukihara et al. (1999). The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. Science 272, 1136-1144.</td>
<td></td>
</tr>
</tbody>
</table>

[0825] XVIII. Biosensors and Environmental Applications
[0826] XVIII.A. Minicell-Based Biosensors
[0827] The present invention is directed to a device that comprises a sensor adapted to detect one or more specific health and/or nutrition markers in a subject or in the environment. The device may also signal the examiner, the subject, or an operator of the occurrence. The sensor comprises a biosensor. As used herein, the term “biosensor” is defined as a component comprising one or more binding motives being adapted to detect a ligand found in one or more target pathogenic microorganisms or related biomolecules.

[0828] Generally, biosensors function by providing a means of specifically binding, and therefore detecting, a target biologically active analyte. In this way, the biosensor is highly selective, even when presented with a mixture of many chemical and biological entities. Often the target biological analyte is a minor component of a complex mixture comprising a multiplicity of biological and other components. Thus, in many biosensor applications, detection of target analytes occurs at parts-per-billion, parts-per-trillion, or even lower numbers.

[0829] XVIII.A.2. Minicell-Based Biosensor Design
[0830] The biosensor of the present invention may comprise a bio-recognition element, or molecular recognition element, that provides the highly specific binding or detection selectivity for a particular analyte. In a biosensor of the invention, the bio-recognition element, or system, is a minicell displaying an enzyme or sequence of enzymes; an antibody or antibody derivative; a membrane receptor protein; or the like, and generally functions to interact specifically with a target biological analyte. The bio-recognition element is responsible for the selective recognition of the analyte and the physico-chemical signal that provides the basis for the output signal. The expressed protein or molecule does not need to be a naturally occurring membrane bound protein but could be a soluble protein or small molecule tethered to the minicell by, for example, a transmembrane domain of another protein such as the EGFR or ToxR.

[0831] Biosensors may include biocatalytic biosensors, and bioaffinity biosensors. In biocatalytic biosensor embodiments, the bio-recognition element minicell is “biocatalytic,” e.g., displays an enzyme. In biocatalytic biosensors, the selective binding sites “turn over” (i.e., can be used again during the detection process), resulting in a significant amplification of the input signal. Biocatalytic sensors such as these are generally useful for real-time, continuous sensing.

[0832] Bioaffinity sensors are generally applicable to bacteria, viruses, toxins and other undesirable compounds and
include chemoreceptor-based biosensors and/or immunological sensors (i.e., immunosensors). Chemoreceptors are complex biomolecular macromolecules responsible, in part, for a viable organism's ability to sense chemicals in its environment with high selectivity. Chemoreceptor-based sensors comprise one or more natural or synthetic chemoreceptors associated with a means to provide a signal (visual, electrical, etc.) of the presence or concentration of a target biological analyte. In certain embodiments, the chemoreceptor may be associated with an electrode (i.e., an electrical transducer) so as to provide a detectable electrical signal. In the biosensors of the invention, minicells displaying a receptor are used in place of chemoreceptors. The minicell has many desired features of a viable cell, and performs similar functions, but is more durable.

[0833] On the other hand, the bio-recognition elements of immunosensors are generally antibodies or antibody derivatives. In any case, bioaffinity biosensors are generally irreversible because the receptor sites of the biosensor become saturated when exposed to the target biological analyte. In a biosensor of the invention, an immunosensor may be a minicell displaying an antibody or antigen fragment.


[0835] The biosensors of the present invention may detect biologically active analytes related to impending (i.e., future presentation of symptoms is likely) or current human systemic disease states, including, but not limited to, pathogenic bacteria, parasites (e.g., any stage of the life cycle, including eggs or portions thereof, cysts, or mature organisms), viruses, fungi such as Candida albicans, antibodies to pathogens, and/or microbially produced toxins. Additionally, the biosensor may target biologically active analytes related to impending or current localized health issues, such as stress proteins (e.g., cytokines) and interleukin 1-alpha that may precede the clinical presentation of skin irritation or inflammation. In preferred embodiments, the biosensor functions as a proactive sensor, detecting and signaling the subject, a caretaker or medical personnel of the impending condition prior to the presentation of clinical symptoms. This allows time to administer prophylactic or remedial treatments to the subject which can significantly reduce, if not prevent, the severity and duration of the symptoms. Further, the sensor, by detecting the presence of a target biological analyte in a sample from the subject, may detect residual contamination on a surface, such as skin or environmental surface, in contact with the biosensor, and provide appropriate signal.

[0836] The physico-chemical signal generated by the bio-recognition element or elements may be communicated visually to the caretaker or medical personnel (i.e., via a color change visible to the human eye). Other embodiments may produce optical signals, which may require other instrumentation to enhance the signal. These include fluorescence, bioluminescence, total internal reflectance resonance, surface plasmon resonance, Raman methods and other laser-based methods, such as LED or laser diode sensors. For example, exemplary surface plasmon resonance biosensors are available as IBIS I and IBIS II from XanTec Analysysysteme of Muenster, Germany, which may comprise bioconjugate surfaces as bio-recognition elements. Alternatively, the signal may be processed via an associated transducer which, for example, may produce an electrical signal (e.g., current, potential, inductance, or impedance) that may be displayed (e.g., on a readout such as an LED or LCD display) or which triggers an audible or tactile (e.g., vibration) signal or which may trigger an actuator, as described herein. The signal may be qualitative (e.g., indicating the presence of the target biological analyte or quantitative (i.e., a measurement of the amount or concentration of the target biological analyte). In such embodiments, the transducer may optionally produce an optical, thermal or acoustic signal.

[0837] In any case, the signal may also be durable (i.e., stable and readable over a length of time typically at least of the same magnitude as the usage life of the device) or transient (i.e., registering a real-time measurement). Additionally, the signal may be transmitted to a remote indicator site (e.g., via a wire, or transmitter, such as an infrared or rf transmitter) including other locations within or on the device or remote devices. Further, the sensor, or any of its components, may be adapted to detect and/or signal only concentrations of the target biological analyte above a predefined threshold level (e.g., in cases wherein the target biological analyte is normally present in the bodily waste or when the concentration of the analyte is below a known "danger" level).

[0838] The target analytes of the biosensors of the present invention are adapted to detect may also be viruses. These may include diarrhea-inducing viruses such as rotavirus, or other viruses such as rhinovirus and human immunodeficiency virus (HIV). An exemplary biosensor adapted to detect HIV is described in U.S. Pat. Nos. 5,830,341 and 5,795,453, referenced above. The disclosure of each of these patents is incorporated by reference herein. Biosensors are adapted to use in different tissues; see e.g., U.S. Pat. No. 6,342,037; Roe et al. Jan. 29, 2002; Device having fecal component sensor; and using different binding molecules, see e.g., U.S. Pat. No. 6,329,160; Schneider et al. Dec. 11, 2001; Biosensors.

[0839] When minicells are incorporated into a biosensor, they may be immobilized in the biosensor by techniques known in the art such as entrapment, adsorption, crosslinking, encapsulation, covalent attachment, any combination thereof, or the like. Further, the immobilization can be carried out on many different substrates such as known in the art. In certain preferred embodiments, the immobilization substrate may be selected from the group of polymer-based materials, hydrogels, tissues, nonwoven materials or woven materials.

[0840] In certain embodiments, biosensor embodiments, may comprise, be disposed on, or be operatively associated with a microchip, such as a silicon chip, MEMS (i.e., micro electromechanical system) device, or an integrated circuit. Microchip-based biosensors may be known as “biochips”. Regardless of the type of sensor, the microchip may comprise a multiplicity of sensor components having similar or different sensitivities, kinetics, and/or target analytes (i.e., markers) in an array adapted to detect differing levels or combinations of the analyte(s). Further, each sensor in such an array may provide a different type of signal, including those types disclosed herein, and may be associated with different actuators and/or controllers. Also, each sensor in an array may operate independently or in association with (e.g., in parallel, combination, or series) any number of other sensors in the array.

[0841] A minicell of a biosensor of the invention may comprise a detectable compound that produces a signal once
ligands have bound to the minicell. By way of non-limiting example, a minicell may display a receptor for a ligand and contain a fluorescent compound. The binding and internalization of the ligand into the minicell results in FRET, shifting the wavelength of the signal. See, by way of non-limiting example, Billington et al., Development of a green fluorescent protein reporter for a yeast genotoxicity biosensor, Biosensors & Bioelectronics 13:831-838, 1998. A biosensor according to the invention may use microbalance sensor systems (Hengen et al., Determination of phage antibody affinities to antigen by a microbalance sensor system, Biotechniques 26:956-964, 1999).

[0842] XVIII.A.2. Surface Plasmon Resonance

[0843] Kd is measured using surface plasmon resonance on a chip, for example, with a BIAcore® chip coated with immobilized binding components, or similar systems such as the IAsys from Thermo Labsystems, Affinity Sensors Division (Cambridge, U.K.) or the BIOS-1 system from Artificial Sensing, Inc. (Zurich, Switzerland). See Fitzgerald, Coupling optical biosensor technology with micropreparative HPLC: Part 1, Am Biotech Lab November 2000, p. 10 and 12; Fitzgerald, Coupling optical biosensor technology with micropreparative HPLC: Part 2, Am Biotech Lab February 2001, 14, 16 and 18; and Leatherbarrow et al., Analysis of molecule recognition using optical sensors, Current Opinion in Chem Biol 3:544-547, 1999).


[0845] BIAcore® uses the optical properties of surface plasmon resonance (SPR) to detect alterations in protein concentration bound within a dextran matrix lying on the surface of a gold/glass sensor chip interface, a dextran biosensor matrix. In brief, proteins are covalently bound to the dextran matrix at a known concentration and a ligand for the protein (e.g., antibody) is injected through the dextran matrix. Near infrared light, directed onto the opposite side of the sensor chip surface is reflected and also induces an evanescent wave in the gold film, which in turn, causes a density dip in the reflected light at a particular angle known as the resonance angle. If the refractive index of the sensor chip surface is altered (e.g., by ligand binding to the bound protein) a shift occurs in the resonance angle. This angle shift can be measured and is expressed as resonance units (RUs) such that 1000 RUs is equivalent to a change in surface protein concentration of 1 ng/mm². These changes are displayed with respect to time along the y-axis of a sensorgram, which depicts the association and dissociation of any biological reaction.


[0847] XVIII.B. Toxicological Sampling

[0848] Minicells are ideally suited for in vitro diagnostic toxicological applications in which toxins, poisons, infectious agents or pathogens, heavy metals, pollutants, caustic agents, allergens, organic molecules, radionuclides, or other environmental contaminants present either in air, water, soil samples and/or fluid and/or tissue samples of organisms can be assessed. An embodiment of this invention, minicells expressing proteins or other molecules could be used in variety of diagnostic detection platforms, including microwell formats, lateral flow devices, molecular switches, biosensors, badges and other sensing devices. Without being limited to the following examples, such devices could be used for early warning of chemical and/or bioweapon attack, illegal drug detection, explosives detection, biohazard detection, pollution assessment, pesticide contamination, allergen detection and detection of toxic or hazardous gasses. In a related application, minicells could be used to eliminate, modify or inactivate the agents.

[0849] In one non-limiting example of protein expression on minicells for toxicological detection, olfactory receptors could be expressed by minicells. The olfactory system possesses the ability to recognize and differentiate between a wide range of odorants based on odor molecules interacting with specific receptor proteins in the ciliary membrane of olfactory neurons (Lancet, 1, 1886. Vertebrate olfactory reception. Ann. Rev. Neurosci. 9:329-355; Shepherd, G. M.,
1994. Discrimination of molecular signals by the olfactory receptor neuron. Neuron 13:771-790). These receptors were found to be 7-transmembrane-domain members of the G protein-coupled receptor family (Buck, L. and R. Alex. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. Cell 65:175-187). Using a murine receptor library, olfactory receptors were functionally expressed in HEK-293 cells (Krautwurst, D., et al., 1998. Identification of ligands for olfactory receptors by functional expression of a receptor library. Cell. 95:917-926). By coexpressing the cloned receptors with Gβ15,16 subunits, the modified receptor system upon activation leads to an increase in intracellular Ca²⁺. Calcium levels were measured employing the dye FURA-2 and ratio fluorimetric imaging. This system demonstrated ligand specificity and structure-function relationships for identified olfactory receptors. Employing similar techniques, OR17-40, a human olfactory receptor protein, was expressed in human embryonic kidney 293 cells and Xenopus Laevis oocytes (Wetzell, H., et al. 1999. Specificity and sensitivity of a human olfactory receptor functionally expressed in Human Embryonic Kidney 293 Cells and Xenopus Laevis Oocytes. J. Neurosciences. 19:7426-7433). The receptor was functionally expressed in a manner designed to assess the specificity of its binding to the ligand, helical.

[0850] In one non-limiting example of target protein identification, primers from homologous areas in transmembrane II and transmembrane VII of olfactory GPCRs will be used to identify unique receptor sequences. These sequences are inserted into expression vectors. Minicell producing bacteria are transformed with these vectors and cultured. Minicells are isolated from the culture as previously described and subsequently induced. Using HTS previously described, the functional receptor/minicell which generates signal for binding of an odorous toxin to the receptor are identified. Large scale-Large scale production of the minicells is carried out and the minicells are covalently coupled to the surface of a microarray chip. The chip is supported in an air sampler, which feeds atmosphere over surface of the chip on a continuous basis. If the toxic agent is present in the air, the binding to the receptor activates a series of events ending in the generation of a signal identifying the presence of the agent in the air.

[0851] By way of non-limiting example, standard molecular biological techniques can be used as follows: cDNA for GFP is ligated into the S end of cDNA sequence for the receptor described above. The resulting sequence is inserted into an expression vector. Minicell producing bacteria are transformed with these vectors and cultured. Minicells are isolated from the culture as previously described and subsequently induced. The minicells now contain the receptor to the ligand on the surface of the minicell with a GFP tag on the C-terminus of the protein in the cytosol. These minicells are packed into filters. Air is passed through the filter. If the ligand is present, it will bind to the receptor. The filter packing is suspended on applied to a diagnostic device. Antibody to the ligand/receptor binding site complex is fixed on the capture zone. When the sample is applied to the device, the receptor/ligand complex is captured. The capture zone is screened for signal resulting from the presence of GFP. This can be extrapolated to have multiple unique receptor/minicell moi eties in the same sampling device. Each receptor would have a unique fluorescing protein tag such that different emissions identify specific agents in the air.

[0852] Other methods for quantification associated take advantage of the composition of the minicell. Loading of the minicell by transiently permeabilizing the membrane to allow for migration of molecules into to the cytosol. These molecules include but are not limited to radiolabeled molecules (i.e., nucleotides), stains or dyes (DAPI) or other DNA staining, heavy metals, fluorophores. The molecules could also be synthesized within the minicell (i.e. GFP). The association of a specific ligand with the minicell could cause a redox shift that induce a color change in the solution or could shift the energy potential in the reaction are generating an electrical current. Each of these examples are associated with well known methods for measuring each of the resulting changes. These include but are not limited to radioactivity or fluorescence generated or the color shift by spectrophotometry.

[0853] A multigene family of gustatory G protein-coupled receptors expressed in the lingual epithelium has been identified with structural similarities to olfactory receptors (Abe, K., et al. 1993. Multiple genes for G protein-coupled receptors and their expression in lingual epithelia. FEBS. 316:253-256; Abe, K., et al. 1993. Primary structure and cell-type specific expression of a gustatory G protein-coupled receptor related to olfactory receptors. J. Bio. Chem.). This provides an addition example of receptors which can be isolated, expressed in minicells and then be used for identification of specific substances in various matrices in similar manners as identified for olfactory receptor minicells.

[0854] As a non-limiting example of minicell use in toxicological/environmental detection, arrays could be constructed in which each well contains a distinct minicell sub-type displaying membrane-bound proteins or other molecules for each of several potential toxins or agents in the environment. For example, minicells in such a format could be used to determine which agents are present in the environment as a consequence of a chemical and/or biological weapons attack. Non-limiting examples of biosensors that have been used toxicological/environmental detection include those described by Sticher et al., Development and characterization of a whole-cell bioluminescent sensor for bioavailable middle-chain alkanes in contaminated groundwater samples, Appl. Eniv. Microbiol. 63:4053-4060, 1997; Willardson et al., Development and Testing of a Bacterial BioSensor for Toluene-Based Environmental Contaminants, Appl. Eniv. Microbiol. 64:1006-1012, 1998; Lars et al., Detection of Oxytetracycline Production by Streptomyces rimosus in Soil Microcosms by Combining Whole-Cell Biosensors and Flow Cytometry, Appl. Eniv. Microbiol. 67:239-244, 2001; Hojberg et al., Oxygen-Sensing Reporter Strain of Pseudomonas fluorescens for Monitoring the Distribution of Low-Oxygen Habitats in Soil, Appl. Eniv. Microbiol. 1999 65: 4085-4093, 1999; R. P. Hollis et al., Design and Application of a BioSensor for Monitoring Toxicity of Compounds to Eukaryotes, Appl. Eniv. Microbiol. 66: 1676-1679, 2000; Heitzer et al., Optical biosensor for environmental on-line monitoring of naphthalene and salicylate bioavailability with an immobilized bioluminescent catalytic reporter bacterium, Appl. Eniv. Microbiol. 60:1487-1494, 1994; Selifonova et al., Bioluminescent sensors for detection of bioavailable Hg(II) in the environment, Appl. Eniv. Microbiol. 59: 3083-3090, 1993; Jaeger et al., Mapping of Sugar and Amino Acid Availability in Soil around Roots with Bacterial Sensors of Sucrose and Tryptophan, Appl. Eniv. Microbiol. 65: 2685-2690, 1999; and Larsen et al., A Microsensor for Nitrate

[0855] XVIII.C. Toxin Elimination

[0856] In another embodiment of the invention, minicells displaying a receptor for a particular toxin agent could be used for the elimination of the agent from the environment. In a non-limiting example of this technology, minicells could be placed in a filtering apparatus to eliminate the toxic agent from the environment (e.g., air, water, soil). In the example of atmospheric contamination, the air would be circulated through a forced air system containing in-line filters composed of a housing, support matrix and receptor/minicells. As air passes over the minicells, the toxin is bound to the receptor.

[0857] Representative toxins include, but are not limited to those associated with “red tides”; bacterial toxins, such as those toxins produced by *Corynebacterium diphteriae* (diphtheria), *Bordetella pertussis* (whooping cough), *Vibrio cholerae* (cholera), *Bacillus anthracis* (anthrax), *Clostridium botulinum* (botulism), *Clostridium tetani* (tetanus), and enterohemorrhagic *Escherichia coli* (bloody diarrhea and hemolytic uremic syndrome); and fungal toxins (e.g., aflatoxin, gliotoxin, cyclopentades, orellanine, gyromitrin, coprine, muscarine, ibotenic acid, psilocybin, psilocin and baecystin).

[0858] The treatment of “red tides” with minicells exemplifies this aspect of the invention. A red tide occurs as a result of a higher-than-normal concentration of an algae or dinoflagellate which, when present in dense concentrations as a result of a “bloom,” form colored patches on the surface of water. The colored patches are pink, violet, orange, yellow, blue, green, brown or red, with red being the most common color. The organisms that cause red tides often produce toxins that have negative impacts on other organisms, including humans.

[0859] For example, *Karenia brevis* (formerly *Gymnodinium breve*) produces a toxin (domoic acid) that affects the central nervous system of fish, shellfish and other organisms, resulting in a state of paralysis. *Alexandrium* species (e.g., *A. tamarense*, *A. fundyense*), *Dinophysis* and *Gonyaulax* species; and *Pseudo-nitzschia multiseries*, which cause, respectively, paralytic, diarrhetic and amnesic shellfish poisoning. Because shellfish containing the toxin taste and appear the same as shellfish that do not, and cooking does not destroy the toxin, human ingestion of the former can cause disease in humans and other organisms. For example, one form of paralytic shellfish poisoning, which can be fatal to humans, results from saxitoxin, which is produced by *Gonyaulax tamarense*, *Protogonyaulax catenella*, and other species. Other algae that can result in red tides include *Gonyaulax catenella*, and *Pychodiscus breve*.

[0860] Minicells that comprise a binding moiety of an organism that produces a red tide, or of the toxin produced thereby, can be used for remediation. For example, a minicell having a binding moiety directed to a red tide-producing organism can be used to deliver an antibiotic thereto, and a minicell with a binding moiety directed to a toxin can be used to bind and/or internalize the toxin. As is explained in more detail elsewhere herein, a minicell with a binding moiety directed to a toxin can also be used for therapeutic purposes.

[0861] XVIII.D. Bioremediation

[0862] In another non-limiting example of the potential use of minicells in a toxicological context is their use in bioremediation, the process by which living organisms act to degrade or transform hazardous organic contaminants. As used herein, “bioremediation” is the process of using biological or biologically derived compositions that alter the chemical structure and/or bind, an undesirable substance in order to reduce the effective concentration of the undesirable substance, thereby reducing or eliminating the effect(s) of the undesirable compound on the environment. Undesirable substances include, but are not limited to pollutants (e.g., heavy metals, pesticides, herbicides, petroleum products); biological toxins (e.g., those produced by “red tides”, e.g., domoic acid, saxitoxin); pathogens (e.g., viruses, eu-bacteria); organisms that produce toxins; biological waste products (e.g., sewage, guano), and undesirable organisms therewithin (e.g., pathogenic eu-bacteria).

[0863] The term “bioremediation” encompasses both biodegradation, the breakdown of organic substances by microorganisms, and biotransformation, the alteration of the structure of a compound by a living organism or enzyme. The minicells of the invention may be incorporated into biofilters, i.e., devices in which gases, liquids, powders and the like are passed through media containing biodegrading minicells, including but not limited to devices that biodegrade volatile organic compounds in air by passing the air therethrough.

[0864] Bioremediation can be used to process undesirable substances in a composition prior to or after the release of the composition into the environment. For example, bioremediation can be applied in sewage treatment plants to process sewage prior to its release, or to sewage that has been accidentally or otherwise released into the environment.

[0865] Environmental microbiologists have sought to identify and use specific bacteria that degrade pollutants and other environmental contaminants. See, for example, Chakrabarty, Microbial Degradation of Toxic Chemicals: Evolutionary Insights and Practical Considerations, Am. Soc. Micro. Biol. News 62:130-137, 1996; and U.S. Pat. Nos. 4,511,657; 4,493,895; 4,871,673; and 4,535,061. In instances where a live organism is placed into the environment to process undesirable substances, there is a concern that the organism might have undesirable effects that would be made more deleterious due to the ability of the live organism to replicate (Saylor G S, Ripp S. Field applications of genetically engineered microorganisms for bioremediation processes. Curr Opin Biotechnol. 2000 June; 11(3):286-9; and Diaz E, Fernandez A, Prieto M A, Garcia J L. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol Mol Biol Rev. 2001 December; 65(4):523-69). For example, when it has been proposed to use genetically altered eu-bacteria to process oil spills, the concern has been raised that the eu-bacteria might spread beyond the oil spill and into supplies of petroleum products that are used to produce energy, where they would process and render useless the stored petroleum products. However, because they
lack the ability to replicate, such a scenario will not occur when minicells are use for bioremediation.

By way of non-limiting example, octane enhances such as methyl t-butyl ether or aromatic hydrocarbons contaminate the aquifer and soil. These agents negatively impact the many microbes in the affected area thus limiting capability of the microbial community rectify the environmental insult. Bioaugmentation, the addition to the environment of microorganisms that can metabolize and grow on specific organic compounds, to facilitate degradation may prove useful, but concerns exist relative to the regulation of newly introduced bacteria. The minicell provides a vehicle to accomplish biodegradation without bacterial overgrowth.

Diphenyl ethers and cyclic ethers such as dioxane and furan have shown to be metabolized by soil bacteria. Using classic isolation and screening techniques identified above, genes encoding for the oxygenases or hydrolyases are isolated. The enzyme sequence is inserted into an expression vector using standard molecular biology techniques. Minicell producing bacteria are transformed with these vectors and cultured. Minicells are isolated from the culture as previously described and subsequently induced. The minicells are applied to the area contaminated with aromatic hydrocarbons. These compounds are transported either actively or passively in to the minicell and subsequently degraded by the oxygenase or hydrolyase. One advantage of this focused degradation is the minimizing of feedback inhibition because the only machinery of consequence in the minicell is that related to the degradation of the ether compounds.

Similarly, beginning with genetic material from Dehalobacter enzymes responsible for the biodegrading of tetrachloroethane could be isolated as described above. The sequence for the enzyme is inserted into the expression vector and used to transform minicell-producing bacteria. The bacteria are cultured, minicells isolated from the culture and the minicells induced as previously described. Minicell preps are lyophilized using standard lyophilization techniques. The resulting material is transported to the site of tetrachloroethene contamination and reconstituted and applied. As the tetrachloroethene was assimilated, it is degraded by the enzyme system.

These are non-limiting examples scope of bioremediation/biotransformation using minicell technology. The scope of the invention includes taking advantage of metabolic pathways organism in general to include but not limited to eukaryotes, prokaryotes, fungi, animals or plants.

Fermentation

Delivery of specific enzymes in an untargeted fashion by the minicell allows for packaged delivery without the increased biomass and complex metabolic products associated with processes using live organisms. This aspect can be taken advantage of in fermentation, where the addition of minicells into which unique enzymes have been added are used to modulate the composition of the environment to include but not limited to the alcohol, sugar and acid levels.

Pesticides

Bacillus thuringensis produces a toxin that kills plant chewing insect larvae as well as mosquito larvae. The toxin, Cry1Ac, binds to aminopeptidase N receptor on the endothoelium of the midgut. Minicell technology is allows for delivery of the toxin. The toxin sequence is modified by ligation of a sequence coding for a transmembrane domain as previously described. The sequence for this fusion protein inserted into an expression vector using standard molecular biology techniques. To facilitate the consumption of the toxin/minicell plasmids containing sequences incorporating the sequence for pheromones coupled at the C-terminus to the sequence for a transmembrane domain is generated using standard molecular biological techniques. This fusion protein sequence is inserted into the expression containing coding region for the toxin fusion protein or inserted into a unique expression vector. Minicell producing bacteria are transformed with these vectors and cultured. Minicells are isolated from the culture as previously described and subsequently induced. The minicells are distributed (e.g. crop dusting) to the area of infestation. The toxin/minicells are ingested by the larvae and kill the larvae as the minicells passes through the gut.
bilize the biologically active form of the composition or one or more of its components. Salts dissolved in buffered solutions are utilized as diluents in the art. For example, preferred diluents are buffered solutions containing one or more different salts. A preferred buffered solution is phosphate buffered saline (particularly in conjunction with compositions intended for pharmaceutical administration), as it mimics the salt conditions of human blood. Since buffer salts can control the pH of a solution at low concentrations, a buffered diluent rarely modifies the biological activity of a biologically active peptide.

[0878] An “excipient” is any more or less inert substance that can be added to a composition in order to confer a suitable property, for example, a suitable consistency or to form a drug. Suitable excipients and carriers include, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol cellulose preparations such as, for example, maize starch, wheat starch, rice starch, agar, pectin, xanthan gum, guar gum, locust bean gum, hyaluronic acid, casein potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylcellulose, polyacrylate, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents can also be included, such as cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Other suitable excipients and carriers include hydrogels, gelable hydrocolloids, and chitosan. Chitosan microspheres and microcapsules can be used as carriers. See WO 98/52547 (which describes microsphere formulations for targeting compounds to the stomach, the formulations comprising an inner core ( optionally including a gelled hydrocolloid containing one or more active ingredients), a membrane comprised of a water insoluble polymer (e.g., ethylcellulose) to control the release rate of the active ingredient(s), and an outer layer comprising of a bioadhesive cationic polymer, for example, a cationic polysaccharide, a cationic protein, and/or a synthetic cationic polymer; U.S. Pat. No. 4,895,724. Typically, chitosan is cross-linked using a suitable agent, for example, glutaraldehyde, glyoxal, epichlorhydrin, and succinyldehyde. Compositions employing chitosan as a carrier can be formulated into a variety of dosage forms, including pills, tablets, microparticles, and microspheres, including those providing for controlled release of the active ingredient(s). Other suitable bioadhesive cationic polymers include acidic gelatin, polycarboxylic acid, polyamino acids such as polylysine, polyhistidine, polyornithine, polyanomeric compounds, prolamine, polyamine, diethylaminoethyl dextran (DEAE), DEAE-imine, DEAE-methacrylate, DEAE-acrylamide, DEAE-dextran, DEAE-cellulose, poly-α-amino styrene, polyacrylamide, copolymer, polyvinylpyridine, and poly(ethyleneimine). The compositions of the invention can be formulated in any suitable manner. Minicell compositions may be uniformly (homogenously) or non-uniformly (heterogeneously) dispersed in the carrier. Suitable formulations include dry and liquid formulations. Dry formulations include freeze dried and lyophilized powders, which are particularly well suited for aerosol delivery to the sinuses or lung, or for long term storage followed by reconstitution in a suitable diluent prior to administration. Other preferred dry formulations include those wherein a composition according to the invention is compressed into tablet or pill form suitable for oral administration or compounded into a sustained release formulation. When the composition is intended for oral administration but is to be delivered to epithelium in the intestines, it is preferred that the formulation be encapsulated with an enteric coating to protect the formulation and prevent premature release of the minicell compositions included therein. As those in the art will appreciate, the compositions of the invention can be placed into any suitable dosage form. Pills and tablets represent some of such dosage forms. The compositions can also be encapsulated into any suitable capsule or other coating material. For example, by compression, dipping, pan coating, spray drying, etc. Suitable capsules include those made from gelatin and starch. In turn, such capsules can be coated with one or more additional materials, for example, and enteric coating, if desired. Liquid formulations include aqueous formulations, gels, and emulsions.

[0880] Some preferred embodiments concern compositions that comprise a bioadhesive, preferably a mucoadhesive, coating. A “bioadhesive coating” is a coating that allows a substance (e.g., a minicell composition) to adhere to a biological surface or substance better than occurs absent the coating. A “mucoadhesive coating” is a preferred bioadhesive coating that allows a substance, for example, a composition according to the invention, to adhere better to mucosa occurs absent the coating. For example, micronized particles (e.g., particles having a mean diameter of about 5, 10, 25, 50, or 100 μm) can be coated with a mucoadhesive. The coated particles can then be assembled into a dosage form suitable for delivery to an organism. Preferably, and depending upon the location where the cell surface transport moiety is targeted is expressed, the dosage form is then coated with another coating to protect the formulation until it reaches the desired location, where the mucoadhesive enables the formulation to be retained while the composition interacts with the target cell surface transport moiety.

[0881] The compositions of the invention may be administered to any organism, preferably an animal, preferably a mammal, bird, fish, insect, or arachnid. Preferred mammals include bovine, canine, equine, feline, ovine, and porcine animals, and non-human primates. Humans are particularly preferred. Multiple techniques of administering or delivering a compound exist in the art including, but not limited to, oral, rectal (e.g., an enema or suppository), aerosol (e.g., for nasal or pulmonary delivery), parenteral, and topical administration. Preferably, sufficient quantities of the biologically active peptide are delivered to achieve the intended effect. The particular amount of composition to be delivered will depend on many factors, including the effect to be achieved, the type of organism to which the composition is delivered, delivery route, dosage regimen, and the age, health, and sex of the organism. As such, the particular dosage of a composition incorporated into a given formulation is left to the ordinarily skilled artisan’s discretion.

[0882] Those skilled in the art will appreciate that when the compositions of the present invention are administered as agents to achieve a particular desired biological result, which may include a therapeutic or protective effect(s) (including vaccination), it may be necessary to combine the fusion proteins of the invention with a suitable pharmaceutical carrier. The choice of pharmaceutical carrier and the preparation of the fusion protein as a therapeutic or protective agent will depend on the intended use and mode of administration. Suitable formulations and methods of administration of therapeutically active agents include those for oral, pulmonary, nasal, buccal, ocular, dermal, rectal, or vaginal delivery.
Depending on the mode of delivery employed, the context-dependent functional entity can be delivered in a variety of pharmaceutically acceptable forms. For example, the context-dependent functional entity can be delivered in the form of a solid, solution, emulsion, dispersion, micelle, liposome, and the like, incorporated into a pill, capsule, tablet, suppository, aerosol, droplet, or spray. Pills, tablets, suppositories, aerosols, powders, droplets, and sprays may have complex, multilayer structures and have a large range of sizes. Aerosols, powders, droplets, and sprays may range from small (1 micron) to large (200 micron) in size.

Pharmaceutical compositions of the present invention can be used in the form of a solid, a lyophilized powder, a solution, an emulsion, a dispersion, a micelle, a liposome, and the like, wherein the resulting composition contains one or more of the compounds of the present invention, as an active ingredient, in admixture with an organic or inorganic carrier or excipient suitable for enteral or parenteral applications. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use. The carriers which can be used include glucose, lactose, mannose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, medium chain length triglycerides, dextrose, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquid form. In addition auxiliary, stabilizing, thickening and coloring agents and perfumes may be used. Examples of a stabilizing dry agent includes triulose, preferably at concentrations of 0.1% or greater (See, e.g., U.S. Pat. No. 5,314,695). The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the process or condition of diseases.

XX. Small Molecules

The term “small molecule” includes any chemical or other moiety that can act to affect biological processes. Small molecules can include any number of therapeutic agents presently known and used, or can be small molecules synthesized in a library of such molecules for the purpose of screening for biological function(s). Small molecules are distinguished from macromolecules by size. The small molecules of this invention usually have molecular weight less than about 5,000 daltons (Da), preferably less than about 2,500 Da, more preferably less than about 1,000 Da, most preferably less than about 500 Da.

Small molecules include without limitation organic compounds, peptidomimetics and conjugates thereof. As used herein, the term “organic compound” refers to any carbon-based compound other than macromolecules such as nucleic acids and polypeptides. In addition to carbon, organic compounds may contain calcium, chlorine, fluorine, copper, hydrogen, iron, potassium, nitrogen, oxygen, sulfur and other elements. An organic compound may be in an aromatic or aliphatic form. Non-limiting examples of organic compounds include acetones, alcohols, alkanes, carbohydrates, monosaccharides, oligosaccharides, polysaccharides, amino acids, nucleosides, nucleotides, lipids, retinoids, steroids, proteoglycans, ketones, aldehydes, saturated, unsaturated and polyunsaturated fats, oils and waxes, alkenes, esters, ethers, thiols, sulfides, cyclic compounds, heterocyclic compounds, imidazoles and phenols. An organic compound as used herein also includes nitrate organic compounds and halogenated (e.g., chlorinated) organic compounds. Methods for preparing peptidomimetics are described below. Collections of small molecules, and small molecules identified according to the invention are characterized by techniques such as accelerator mass spectrometry (AMS; see Turteltaub et al., Curr Pharm Des 2006 12(10):991-1007, Bioanalytical applications of accelerator mass spectrometry for pharmaceutical research; and Enjalbal et al., Mass Spectrom Rev 2000 19(3):139-61, Mass spectrometry in combinatorial chemistry.)

Preferably small molecules are relatively easier and less expensively manufactured, formulated or otherwise prepared. Preferably small molecules are stable under a variety of storage conditions. Preferably small molecules may be placed in tight association with macromolecules to form molecules that are biologically active and that have improved pharmaceutical properties. Improved pharmaceutical properties include changes in circulation time, distribution, metabolism, modification, excretion, secretion, elimination, and stability that are favorable to the desired biological activity. Improved pharmaceutical properties include changes in the toxicological and efficacy characteristics of the chemical entity.

XXI. Polypeptides and Derivatives
XXI.A. Polypeptides

As used herein, the term “polypeptide” includes proteins, fusion proteins, oligopeptides and polypeptide derivatives, with the exception that peptidomimetics are considered to be small molecules herein. Although they are polypeptides, antibodies and their derivatives are described in a separate section. Antibodies and antibody derivatives are described in a separate section, but antibodies and antibody derivatives are, for purposes of the invention, treated as a subclass of the polypeptides and derivatives.

A “protein” is a molecule having a sequence of amino acids that are linked to each other in a linear molecule by peptide bonds. The term protein refers to a polypeptide that is isolated from a natural source, or produced from an isolated cDNA using recombinant DNA technology, and has a sequence of amino acids having a length of at least about 200 amino acids.

A “fusion protein” is a type of recombinant protein that has an amino acid sequence that results from the linkage of the amino acid sequences of two or more normally separate polypeptides.

A “protein fragment” is a proteolytic fragment of a larger polypeptide, which may be a protein or a fusion protein. A proteolytic fragment may be prepared by in vivo or in vitro proteolytic cleavage of a larger polypeptide, and is generally too large to be prepared by chemical synthesis. Proteolytic fragments have amino acid sequences having a length from about 200 to about 1,000 amino acids.

An “oligopeptide” is a polypeptide having a short amino acid sequence (i.e., 2 to about 200 amino acids). An oligopeptide is generally prepared by chemical synthesis.

Although oligopeptides and protein fragments may be otherwise prepared, it is possible to use recombinant DNA technology and/or in vitro biochemical manipulations. For example, a nucleic acid encoding an amino acid sequence may be prepared and used as a template for in vitro transcription/translation reactions. In such reactions, an exogenous nucleic acid encoding a preselected polypeptide is introduced into a mixture that is essentially depleted of exogenous nucleic acids that contains all of the cellular components required for transcription and translation. One or more radio-
labeled amino acids are added before or with the exogenous DNA, and transcription and translation are allowed to proceed. Because the only nucleic acid present in the reaction mix is the exogenous nucleic acid added to the reaction, only polypeptides encoded thereby are produced, and incorporate the radiolabelled amino acid(s). In this manner, polypeptides encoded by a preselected exogenous nucleic acid are radiolabeled. Although other proteins are present in the reaction mix, the preselected polypeptide is the only one that is produced in the presence of the radiolabeled amino acids and is thus uniquely labeled.

[0897] As is explained in detail below, “polypeptide derivatives” include without limitation mutant polypeptides, chemically modified polypeptides, and peptidomimetics.

[0898] The polypeptides of this invention, including the analogs and other modified variants, may generally be prepared following known techniques. Preferably, synthetic production of the polypeptide of the invention may be according to the solid phase synthetic method. For example, the solid phase synthesis is well understood and is a common method for preparation of polypeptides, as are a variety of modifications of that technique [Merrifield (1964), J. Am. Chem. Soc., 85: 2149; Stewart and Young (1984), Solid Phase polypeptide Synthesis, Pierce Chemical Company, Rockford, Ill.; Bodansky and Bodansky (1984), The Practice of polypeptide Synthesis, Springer-Verlag, New York; Atherton and Sheppard (1989), Solid Phase polypeptide Synthesis: A Practical Approach, IRL Press, New York]. See also, the specific method described in Example 1 below.

[0899] Alternatively, polypeptides of this invention may be prepared in recombinant systems using polynucleotide sequences encoding the polypeptides. For example, fusion proteins are typically prepared using recombinant DNA technology.

[0900] XXI.B. Polypeptide Derivatives

[0901] A “derivative” of a polypeptide is a compound that is not, by definition, a polypeptide, i.e., it contains at least one chemical linkage that is not a peptide bond. Thus, polypeptide derivatives include without limitation proteins that naturally undergo post-translational modifications such as, e.g., glycosylation. It is understood that a polypeptide of the invention may contain more than one of the following modifications within the same polypeptide. Preferred polypeptide derivatives retain a desirable attribute, which may be biological activity; more preferably, a polypeptide derivative is enhanced with regard to one or more desirable attributes, or has one or more desirable attributes not found in the parent polypeptide. Although they are described in this section, peptidomimetics are taken as small molecules in the present disclosure.

[0902] XXI.C. Mutant Polypeptide Derivatives

[0903] A polypeptide having an amino acid sequence identical to that found in a protein prepared from a natural source is a “wildtype” polypeptide. Mutant oligopeptides can be prepared by chemical synthesis, including without limitation combinatorial synthesis.

[0904] Mutant polypeptides larger than oligopeptides can be prepared using recombinant DNA technology by altering the nucleotide sequence of a nucleic acid encoding a polypeptide. Although some alterations in the nucleotide sequence will not alter the amino acid sequence of the polypeptide encoded thereby (“silent” mutations), many will result in a polypeptide having an altered amino acid sequence that is altered relative to the parent sequence. Such altered amino acid sequences may comprise substitutions, deletions and additions of amino acids, with the proviso that such amino acids are naturally occurring amino acids.

[0905] Thus, subjecting a nucleic acid that encodes a polypeptide to mutagenesis is one technique that can be used to prepare mutant polypeptides, particularly ones having substitutions of amino acids but no deletions or insertions thereof. A variety of mutagenic techniques are known that can be used in vitro or in vivo including without limitation chemical mutagenesis and PCR-mediated mutagenesis. Such mutagenesis may be randomly targeted (i.e., mutations may occur anywhere within the nucleic acid) or directed to a section of the nucleic acid that encodes a stretch of amino acids of particular interest. Using such techniques, it is possible to prepare randomized, combinatorial or focused compound libraries, pools and mixtures.

[0906] Polypeptides having deletions or insertions of naturally occurring amino acids may be synthetic oligopeptides that result from the chemical synthesis of amino acid sequences that are based on the amino acid sequence of a parent polypeptide but which have one or more amino acids inserted or deleted relative to the sequence of the parent polypeptide. Insertions and deletions of amino acid residues in polypeptides having longer amino acid sequences may be prepared by directed mutagenesis.

[0907] XXII.D. Chemically Modified Polypeptides

[0908] As contemplated by this invention, the term “polypeptide” includes those having one or more chemical modification relative to another polypeptide, i.e., chemically modified polypeptides. The polypeptide from which a chemically modified polypeptide is derived may be a wildtype protein, a mutant protein or a mutant polypeptide, or polypeptide fragments thereof; an antibody or other polypeptide ligand according to the invention including without limitation single-chain antibodies, bacterial proteins and polypeptide derivatives thereof; or polypeptide ligands prepared according to the disclosure. Preferably, the chemical modification(s) confer(s) or improve(s) desirable attributes of the polypeptide but does not substantially alter or compromise the biological activity thereof. Desirable attributes include but are limited to increased shelf-life; enhanced serum or other in vivo stability; resistance to proteases; and the like. Such modifications include by way of non-limiting example N-terminal acetylation, glycosylation, and biotinylation.

[0909] XXII.D.1. Polypeptides with N-Terminal or C-Terminal Chemical Groups

[0910] An effective approach to confer resistance to proteases acting on the N-terminal or C-terminal residues of a polypeptide is to add chemical groups at the polypeptide termini, such that the modified polypeptide is no longer a substrate for the peptidase. One such chemical modification is glycosylation of the peptidases at either or both termini. Certain chemical modifications, in particular N-terminal glycosylation, have been shown to increase the stability of polypeptides in human serum (Powell et al. (1993), Phar. Res. 10: 1268-1273). Other chemical modifications which enhance serum stability include, but are not limited to, the addition of an N-terminal alkyl group, consisting of a lower alkyl of from 1 to 20 carbons, such as an acetyl group, and/or the addition of a C-terminal amide or substituted amide group.
[0911] XXI.D.2. Polypeptides with a Terminal D-Amino Acid

[0912] The presence of an N-terminal D-amino acid increases the serum stability of a polypeptide that otherwise contains L-amino acids, because exopeptidases acting on the N-terminal residue cannot utilize a D-amino acid as a substrate. Similarly, the presence of a C-terminal D-amino acid also stabilizes a polypeptide, because exopeptidases acting on the C-terminal residue cannot utilize a D-amino acid as a substrate. With the exception of these terminal modifications, the amino acid sequences of polypeptides with N-terminal and/or C-terminal D-amino acids are usually identical to the sequences of the parent L-amino acid polypeptide.


[0914] Substitution of unnatural amino acids for natural amino acids in a subsequence of a polypeptide can confer or enhance desirable attributes including biological activity. Such a substitution can, for example, confer resistance to proteolysis by exopeptidases acting on the N-terminus. The synthesis of polypeptides with unnatural amino acids is routine and known in the art (see, for example, Coller, et al. 1993), cited above).

[0915] XXI.D.4. Post-Translational Chemical Modifications

[0916] Different host cells will contain different post-translational modification mechanisms that may provide particular types of post-translational modification of a fusion protein if the amino acid sequences required for such modifications is present in the fusion protein. A large number (≈100) of post-translational modifications have been described, a few of which are discussed herein. One skilled in the art will be able to choose appropriate host cells, and design chimeric genes that encode protein members comprising the amino acid sequence needed for a particular type of modification.

[0917] Glycosylation is one type of post-translational chemical modification that occurs in many eukaryotic systems, and may influence the activity, stability, pharmacogenetics, immunogenicity and/or antigenicity of proteins. However, specific amino acids must be present at such sites to recruit the appropriate glycosylation machinery, and not all host cells have the appropriate molecular machinery. Saccharomyces cerevisiae and Pichia pastoris provide for the production of glycosylated proteins, as do expression systems that utilize insect cells, although the pattern of glycosylation may vary depending on which host cells are used to produce the fusion protein.

[0918] Another type of post-translation modification is the phosphorylation of a free hydroxyl group of the side chain of one or more Ser, Thr or Tyr residues. Protein kinases catalyze such reactions. Phosphorylation is often reversible due to the action of a protein phosphatase, an enzyme that catalyzes the dephosphorylation of amino acid residues.

[0919] Differences in the chemical structure of amino terminal residues result from different host cells, each of which may have a different chemical version of the methionine residue encoded by a start codon, and these will result in amino termini with different chemical modifications.

[0920] For example, many or most bacterial proteins are synthesized with an amino terminal amino acid that is a modified form of methionine, i.e. N-formyl-methionine (fMet). Although the statement is often made that all bacterial proteins are synthesized with an fMet initiator amino acid, although this may be true for E. coli, recent studies have shown that it is not true in the case of other bacteria such as Pseudomonas aeruginosa (Newton, et al., J. Biol. Chem. 274: 22143-22146, 1999). In any event, in E. coli, the formyl group of fMet is usually enzymatically removed after translation to yield an amino terminal methionine residue, although the entire fMet residue is sometimes removed (see Hershey, Chapter 40, "Protein Synthesis" in: Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology, Neidhardt, Frederick C., Editor in Chief, American Society for Microbiology, Washington, D.C., 1987, Volume 1, pages 613-647, and references cited therein.) E. coli mutants that lack the enzymes (such as, e.g., formylase) that catalyze such post-translational modifications will produce proteins having an amino terminal fMet residue (Guillon, et al., J. Bacteriol. 174:4294-4301, 1992).

[0921] In eukaryotes, acetylation of the initiator methionine residue, or the penultimate residue if the initiator methionine has been removed, typically occurs co- or post-translationally. The acetylation reactions are catalyzed by N-terminal acetyltransferases (NATs), aka N-alpha-acetyltransferases, whereas removal of the initiator methionine residue is catalyzed by methionine aminopeptidases (for reviews, see Bradshaw et al., Trends Biochem. Sci. 23:263-267, 1998; and Driessen et al., CRC Crit. Rev. Biochem. 18:281-325, 1985). Amino terminally acetylated proteins are said to be "N-acetylated," "N alpha acetylated" or simply "acetylated."

[0922] Another post-translational process that occurs in eukaryotes is the alpha-amidation of the carboxy terminus. For reviews, see Epper et al. Annu. Rev. Physiol. 50:333-344, 1988, and Bradbury et al. Lung Cancer 14:239-251, 1996. About 50% of known endocrine and neuroendocrine peptide hormones are alpha-amidated (Treston et al., Cell Growth Differ. 4:911-920, 1993). In most cases, carboxy alpha-amidation is required to activate these peptide hormones.

[0923] XXI.E. Peptidomimetics

[0924] In general, a polypeptide mimetic (“peptidomimetic”) is a molecule that mimics the biological activity of a polypeptide but is no longer peptide in chemical nature. By strict definition, a peptidomimetic is a molecule that contains no peptide bonds (that is, amide bonds between amino acids). However, the term peptidomimetic is sometimes used to describe molecules that are no longer completely peptide in nature, such as pseudo-peptides, semi-peptides and peptoids. Examples of some peptidomimetics by the broader definition (where part of a polypeptide is replaced by a structure lacking peptide bonds) are described below. Whether completely or partially non-peptide, peptidomimetics according to this invention provide a spatial arrangement of reactive chemical moieties that closely resembles the three-dimensional arrangement of active groups in the polypeptide on which the peptidomimetic is based. As a result of this similar active-site geometry, the peptidomimetic has effects on biological systems that are similar to the biological activity of the polypeptide.

[0925] There are several potential advantages for using a mimetic of a given polypeptide rather than the polypeptide itself. For example, polypeptides may exhibit two undesirable attributes, i.e., poor bioavailability and short duration of action. Peptidomimetics are often small enough to be both orally active and to have a long duration of action. There are
also problems associated with stability, storage and immunoactivity for polypeptides that are not experienced with peptidomimetics.

[0926] Candidate, lead and other polypeptides having a desired biological activity can be used in the development of peptidomimetics with similar biological activities. Techniques of developing peptidomimetics from polypeptides are known. Peptide bonds can be replaced by non-peptide bonds that allow the peptidomimetic to adopt a similar structure, and therefore biological activity, to the original polypeptide. Further modifications can also be made by replacing chemical groups of the amino acids with other chemical groups of similar structure. The development of peptidomimetics can be aided by determining the tertiary structure of the original polypeptide, either free or bound to a ligand, by NMR spectroscopy, crystallography and/or computer-aided molecular modeling. These techniques aid in the development of novel compositions of higher potency and/or greater bioavailability and/or greater stability than the original polypeptide (Dean (1994), BioEssays, 16: 683-687; Cohen and Shatzmiller (1993), J. Mol. Graph., 11: 166-173; Wiley and Rich (1993), Med. Res. Rev., 13: 327-354; Moore (1994), Trends Pharmacol. Sci., 15: 124-129; Hruby (1993), Biopolymers, 33: 1073-1082; Bugg et al. (1993), Sci. Am., 269: 92-98, all incorporated herein by reference).

[0927] Thus, through use of the methods described above, the present invention provides compounds exhibiting enhanced therapeutic activity in comparison to the polypeptides described above. The peptidomimetic compounds obtained by the above methods, having the biological activity of the above named polypeptides and similar three-dimensional structure, are encompassed by this invention. It will be readily apparent to one skilled in the art that a peptidomimetic can be generated from any of the modified polypeptides described in the previous section or from a polypeptide bearing more than one of the modifications described from the previous section. It will furthermore be apparent that the peptidomimetics of this invention can be further used for the development of even more potent non-peptidic compounds, in addition to their utility as therapeutic compounds.

[0928] Specific examples of peptidomimetics derived from the polypeptides described in the previous section are presented below. These examples are illustrative and not limiting in terms of the other or additional modifications.

[0929] XXI.E.1. Peptides with a Reduced Isozere Pseudopeptide Bond

[0930] Proteases act on peptide bonds. It therefore follows that substitution of peptide bonds by pseudopeptide bonds confers resistance to proteolysis. A number of pseudopeptide bonds have been described that in general do not affect polypeptide structure and biological activity. The reduced isozere pseudopeptide bond is a suitable pseudopeptide bond that is known to enhance stability to enzymatic cleavage with no or little loss of biological activity (Couder et al. (1993), Int. J. Polypeptide Protein Res. 41:181-184, incorporated herein by reference). Thus, the amino acid sequences of these compounds may be identical to the sequences of their parent L-amino acid polypeptides, except that one or more of the peptide bonds are replaced by an isozere pseudopeptide bond. Preferably the most N-terminal peptide bond is substituted, since such a substitution would confer resistance to proteolysis by exopeptidases acting on the N-terminus.

[0931] XXI.E.2. Peptides with a Retro-Inverso Pseudopeptide Bond

[0932] To confer resistance to proteolysis, peptide bonds may also be substituted by retro-inverso pseudopeptide bonds (Dalpezzo et al. (1993), Int. J. Polypeptide Protein Res. 41:561-566, incorporated herein by reference). According to this modification, the amino acid sequences of the compounds may be identical to the sequences of their L-amino acid parent polypeptides, except that one or more of the peptide bonds are replaced by a retro-inverso pseudopeptide bond. Preferably the most N-terminal peptide bond is substituted, since such a substitution will confer resistance to proteolysis by exopeptidases acting on the N-terminus.

[0933] XXI.E.3. Peptoid Derivatives

[0934] Peptoid derivatives of polypeptides represent another form of modified polypeptides that retain the important structural determinants for biological activity, yet eliminate the peptide bonds, thereby conferring resistance to proteolysis (Simon et al. (1992), Proc. Natl. Acad. Sci. USA, 89:9367-9371 and incorporated herein by reference). Peptoids are oligomers of N-substituted glycines. A number of N-alkyl groups have been described, each corresponding to the side chain of a natural amino acid.

[0935] XXII. Kits

[0936] The invention provides for diagnostic and therapeutic kits related useful for therapeutic, diagnostic, and research applications. Exemplary kits are disclosed in U.S. Pat. Nos. 5,773,024; 6,017,721; and 6,232,127 B1. The kits of the invention incorporate minicells, and/or include methods of using minicells described herein.

[0937] XXII.A. Diagnostic and Research Use Kit Components

[0938] In one embodiment, the invention relates to kits for determining the diagnosis or prognosis of a patient. These kits preferably comprise devices and reagents for measuring one or more marker levels in a test sample from a patient, and instructions for performing the assay. Optionally, the kits may contain one or more means for converting marker level(s) to a diagnosis. Such kits preferably contain sufficient reagents to perform one or more such determinations.

[0939] More specifically, a diagnostic kit of the invention comprises any of the following reagents and/or components in any combination.

[0940] (1) A detectable or detectably labeled first reagent that binds a ligand of interest. The binding reagent can, but need not, be an antibody or an antibody derivative comprising a detectable moiety. The sphingolipid-binding reagent is stored in an openable container in the kit, or is bound to a surface of a substrate such that it is accessible to other reagents. Examples of the latter include test strips.

[0941] (2) If the first reagent in neither detectable nor detectably labeled, the kit may comprise a detectable or detectably labeled second reagent that binds to the first reagent (e.g., a secondary antibody) or which produces a detectable signal when in close proximity to the first reagent (e.g., as results from fluorescent resonance energy transfer FRET). In either case, the signal produced from the second reagent correlates with the amount of ligand in the sample.

[0942] (3) One or more positive control reagents. Typically, these reagents comprise a compound that is known to produce a signal in the assay. In one embodiment, the positive control reagents are standards, i.e., comprise a known amount of a detectable or detectably labeled compound, the signal from which may be compared to the signal from a test sample. In
addition to serving as positive control reagents, they may be used to develop calibration curves that relate the amount of signal to the known concentration of a detectable or detectably labeled compound. The signal from a test sample is compared to the calibration curve in order to determine what concentration of the detectable or detectably labeled compound corresponds to the signal from the test sample. In this embodiment, the kit provides quantitative measurements of the amount of a ligand in a test sample.

(0943) (4) One or more negative control reagents. Typically, these control reagents may comprise buffer or another solution that does not contain any of the detectable or detectably labeled first or second reagents and should thus not produce any detectable signal. Any signal that is detected reflects the background level of "noise" in the assay. Another type of negative control reagent contains most of the components necessary for the signal of the assay to be produced, but lacks at least one such component and therefore should not produce a signal. Yet another type of negative control reagent contains all of the components necessary for the signal of the assay to be produced, but also contains an inhibitor of the process that produced the signal.

(0944) (5) One or more auxiliary reagents for use in the diagnostic assays of the kit, e.g., buffers, alcohols, acid solutions, etc. These reagents are generally available in medical facilities and thus are optional components of the kit. However, these reagents preferably are included in the kit to ensure that reagents of sufficient purity and sterility are used, since the resulting protein conjugates are to be administered to mammals, including humans, for medical purposes, and to provide kits that can be used in situations where medical facilities are not readily available, e.g., when hikers in places located far from medical facilities, or in situations where the presence of these auxiliary reagents allows for the immediate treatment of a patient outside of a medical facility as opposed to treatment that arrives at some later time).

(0945) (6) Instructions to a person using a kit for its use. The instructions can be present on one or more of the kit components, the kit packaging and/or a kit package insert.

XXII.B. Therapeutic Kit Components

(0946) A therapeutic kit of the invention comprises any of the following reagents and/or components in any combination.

(0947) (1) One or more therapeutic agents.

(0948) (2) If the therapeutic agent(s) are not formulated for delivery via the alimentary canal, which includes but is not limited to sublingual delivery, a device capable of delivering the therapeutic agent through some other routes. One type of device for parenteral delivery is a syringe that is used to inject the therapeutic agent into the body of an animal in need of the therapeutic agent. Inhalation devices may also be used.

(0949) (3) Separate containers, each of which comprises one or more reagents of the kit. In a preferred embodiment, the containers are vials contain sterile, lyophilized formulations of a therapeutic composition that are suitable for reconstitution. Other containers include, but are not limited to, a pouch, tray, box, tube, or the like. Kit components may be packaged and maintained sterilely within the containers.

(0950) (4) Instructions to a person using a kit for its use. The instructions can be present on one or more of the kit components, the kit packaging and/or a kit package insert. Such instructions include, by way of non-limiting example, instructions for use of the kit and its reagents, for reconstituting lyophilized reagents or otherwise preparing reagents.

(0951) A preferred kit of the present invention comprises the elements useful for performing an immunosay. A kit of the present invention can comprise one or more experimental samples (i.e., formulations of the present invention) and one or more control samples bound to at least one pre-packed dipstick or ELISA plate, and the necessary means for detecting immunocomplex formation (e.g., labelled secondary antibodies or other binding compounds and any necessary solutions needed to resolve such labels, as described in detail above) between antibodies contained in the bodily fluid of the animal being tested and the proteins bound to the dipstick or ELISA plate. It is within the scope of the invention that the kit can comprise simply a formulation of the present invention and that the detecting means can be provided in another way.

(0952) An alternative preferred kit of the present invention comprises elements useful for performing a skin test. A kit of the present invention can comprise at least one pre-packed syringe and needle apparatus containing one or more experimental samples and/or one or more control samples. A kit according to the invention may be designed for both diagnostic and therapeutic applications. Any combination of the above elements XX.A.(1)-(6) and XX.B.(1)-(4) may be used in a kit, optionally with additional reagents, standards, sample containers, an the like.

(0953) XXIII. Immunogenic Minicells

(0954) XXIII.A. In General

(0955) Minicells are used to immunize subjects. An organism is said to be "immunized" when, after contact with an immunogen, the organism produces antibodies directed to the immunogen, or has increased proliferation or activity of cytotoxic and/or helper T cells, or both. Increased proliferation or activity of T cells may be particularly desirable in the case of parasites that cause a decrease in T cell proliferation.

(0956) The use of minicells to present antigens has several potential advantages. An intact membrane protein can be presented in its native form on the surface of an immunogenic minicell, rather than as a denatured protein or as oligopeptides derived from the amino acid sequence of a membrane protein, which allows for antibodies to be developed that are directed to epitopes which, due to protein folding, occur only in the native protein. The minicell surface may naturally be, or may be modified to be, an adjuvant. Moreover, pharmacokinetic properties of minicells, as discussed elsewhere herein, may be improved relative to other forms of administration.

(0957) The applications of immunogenic minicells include, but are not limited to, research, prophylactic, diagnostic and therapeutic applications.

(0958) In research applications, immunogenic minicells are used to generate antibodies to an antigen displayed on a minicell. Such antibodies are used to detect an antigen, which may be a chemical moiety, molecule, virus, organelle, cell, tissue, organ, or organism that one wishes to study. Classically, such antibodies have been prepared by immunizing an animal, often a rat or a rabbit, and collecting antiserum therefrom. Molecular biology techniques can be used to prepare antibodies and antibody fragments, as is described elsewhere herein. Single-chain antibody fragments (scFv) may also be identified, purified, and characterized using minicells displaying a membrane protein or membrane bound chimeric soluble protein.

(0959) In prophylactic applications, immunogenic minicells are used to stimulate a subject to produce antibodies
and/or activate T cells, so that the subject is “pre-immunized” before contact with a pathogen or hyperproliferative cell. Thus, in the case of a pathogen, the subject is protected by antibodies and/or T cells that are specifically directed to the pathogen before infection.

[0960] In therapeutic applications, immunogenic minicells are used in immunotherapy.

[0961] Certain aspects of the invention involve active immunotherapy, in which treatment relies on the in vivo stimulation of the endogenous host immune system to react against pathogens or tumors due to the administration of agents that cause, enhance or modulate an immune response. Such agents include, but are not limited to, immunogens, adjuvants, cytokines and chemokines.

[0962] Other therapeutic applications involve passive immunotherapy, in which treatment involves the delivery of agents (such as antibodies or effector cells) that are specifically directed to an immunogen of a pathogen or a hyperproliferative cell, and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells; T lymphocytes, such as CD8+ cytotoxic T lymphocytes and CD4+ T-helper tumor-infiltrating lymphocytes; killer cells, such as Natural Killer (NK) cells and lymphokine-activated killer cells.

[0963] XXXIII.B. Hyperproliferative Disorders

[0964] The immunogenic minicells of the invention can be used to treat hyperproliferative disorders by inducing an immune response to an antigen associated therewith. The term “hyperproliferative disorder” refers to disorders characterized by an abnormal or pathological proliferation of cells, for example, cancer, psoriasis, hyperplasia and the like.


[0967] Tumor specific antigens (TSAs), tumor-associated differentiation antigens (TADAs) and other antigens associated with cancers and other hyperproliferative disorders include, but are not limited to, C1 IAC, a human cancer protein (Oster, U.S. Pat. No. 4,132,769); the CA125 antigen, an antigen associated with cystadenocarcinoma of the ovary, (Hanisch et al., Carbohydr. Res. 178:29-47, 1988; O’Brien, U.S. Pat. No. 4,921,790); CEA, an antigen present on many adenocarcinomas (Honi et al., Strategies for cancer therapy using carcinembryonic antigen vaccines, Expert Reviews in Molecular Medicine, http://www-ermm.cbcu.cam.ac.uk: 1, 2000); CORA (carcinoma or orosomucoid-related antigen) described by Toth et al. (U.S. Pat. No. 4,914,021); DF3 antigen from human breast carcinoma (Kufe, in U.S. Pat. Nos. 4,963,484 and 5,053,489); DU-PAN-2, a pancreatic carcinoma antigen (Lan et al., Cancer Res. 45:305-310, 1985); HCA, a human carcinoma antigen (Coldington et al., U.S. Pat. No. 5,693,763); Her2, a breast cancer antigen (Fendly et al., The Extracellular Domain of HER2/ neu is A Potential Immunogen for Active Specific Immunotherapy of Breast Cancer, Journal of Biological Response Modifiers 9:449-455, 1990); MSA, a breast carcinoma glycoprotein (Tjandra et al., Br. J. Surg. 75:811-817, 1988); MFGM, a breast carcinoma antigen (Ishida et al., Tumor Biol. 10:12-22, 1988); PSA, prostate specific antigen (Nadj et al., Prostatic-specific-antigen, Cancer 48:1229-1232, 1981); STEAP (six transmembrane epithelial antigens of the prostate) proteins (Afar et al., U.S. Pat. No. 6,329,503; TAG-72, a breast carcinoma glycoprotein (Kjeldsen et al., Cancer Res. 48:2214-2220, 1988); YH206, a lung carcinoma antigen (Hinoda et al., Cancer J. 42:653-658, 1988); the p97 antigen of human melanoma (Eslin et al., Recombinant Vaccinia Virus Vaccine Against the Human Melanoma Antigen p97 for Use in Immunotherapy, Proc. Natl. Acad. Sci. USA, 85:1052-1056, 1988); and the melanoma specific antigen described by Pfleundschuh in U.S. Pat. No. 6,025,191).

[0968] XXXIII.B. Intracellular Pathogens

[0969] In certain aspects of the invention, vaccines comprising immunogenic minicells are used to prevent or treat diseases caused by intracellular pathogens. Vaccines may be prepared that stimulate cytotoxic T cell responses against cells infected with viruses including, but not limited to, hepatitis type A, hepatitis type B, hepatitis type C, influenza, varicella, adenovirus, herpes simplex type 1 (HSV-1), herpes simplex type 2 (HSV-2), rinderpest, rhinovirus, echovirus, rotavirus, respiratory syncytial virus, papilloma virus, papova virus, cytomegalovirus, echovirus, arbovirus, hantavirus, coxsackie virus, mumps virus, measles virus, rubella virus, polio virus, human immunodeficiency virus type 1 (HIV-1), and human immunodeficiency virus type II (HIV-2). Vaccines also may be prepared that stimulate cytotoxic T cell responses against cells infected with intracellular obligates, including but not limited to Chlamydia, Mycobacteria and Rickettsia. Vaccines also may be prepared that stimulate cytotoxic T cell responses against cells infected with intracellular protozoa, including, but not limited to, leishmania, kokkidi- toa, and trypanosoma.

[0970] The causative agent of Lyme disease, the spirochete Borrelia burgdorferi, is also of interest. The outer surface proteins (Osp) A, B and C of B. burgdorferi are known antigens that are lipoproteins that associate with membranes. Amino-terminal cysteine residues in Osp proteins are the sites of tricarboxylic acid modifications that serve as membrane-anchoring motifs. The N-terminal portions of the Osp proteins are highly conserved and are preferred portions for display on immunogenic minicells.
[0971] XXIII.C. Eukaryotic Pathogens

[0972] In addition to intracellular pathogens, other eukaryotic pathogens exist and may also be treated using immunogenic minicells displaying antigens therefrom. A number of antigens have been developed to combat parasitic vaccines, e.g., the recombinant 45w protein of Taenia ovis; EG95 oncosphere proteins of Echinococcus granulosus; cathepsin L antigen of the liver fluke, Fasciola hepatica; and the H1 antigen of Haemonchus contortus (Dalton et al., Parasite vaccines—a reality?, Vet Parasitol 98:149-167, 2001). Other eukaryotic pathogens include, but are not limited to:

[0973] Protozoans, including but not limited to, Entamoeba histolytica, a pathogenic amoeba that causes amoebic dysentery and occasionally digests its way through the intestinal wall to invade other organs, which may cause morbidity; Balantidium coli, a ciliate that causes diarrhea in humans; Giardia lamblia, a flagellate that causes diarrhea and abdominal pain, along with a chronic fatigue syndrome that is otherwise asymptomatic and difficult to diagnose; Trypanosoma brucei, a hemoflagellate causing sleeping sickness; and Trypanosoma cruzi, the cause of Chagas disease.

[0974] Plasmodia, sporozoites obligate intracellular parasites of liver and red blood cells, including but not limited to P. falciparum, the causative agent of malaria. Dozens of P. falciparum antigens have been identified, e.g., CSP-1, STARP, SALSA, SSP-2, ISA-1, EXP-1, ISA-3, RAP-1, RAP-2, SERA-1, MSP-1, MSP-2, MSP-3, MSP-4, MSP-5, AMA-1, EBA-175, RESA, GLURP, EMP-1, Pf25k, Pf27k, PD35, Pf55, Pf230, Pf27, Pf16, Pf628 and Pf645/48.

[0975] Helminthes including but limited to Ascaris lumbricoides (roundworm); Enterobius vermicularis (pinworm); Trichurus trichiura (whipworm); and Fasciola hepatica (liver fluke);

[0976] Taenia sp. (tapeworms and cestodes);

[0977] Schistosoma (trematodes), such as Schistoma mansoni, which comprises the Sm32 antigen (asparaginyl endopeptidase), which can induce antibody formation in mice (Chichilia et al., DNA vaccination with asparaginyl endopeptidase (Sm32) from the parasite Schistosoma mansoni: anti-fecundity effect induced in mice, Vaccine 20:439-447, 2001); and acetylokininestrase (Arnon et al., Acetylcholinesterase of Schistoma mansoni-Functional correlates, Protein Science 8:2533-2561, 1999); and

[0978] Ticks and other invertebrates, including but not limited to insects, arachnids, etc. For example, a description of a vaccine against the cattle tick Boophilus microplus has been described (Valle et al., The evaluation of yeast derivatives as adjuvants for the immune response to the Bm86 antigen in cattle, BMC Biotechnol. 1:2, 2001)

[0979] XXIII.D. Formulation and Administration of Immunogenic Minicells

[0980] Vaccine formulations of immunogenic minicells include a suitable carrier. Because minicells may be destroyed by digestion, or prevented from acting due to antibody secretion in the gut, they are preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain antioxidants, buffers, and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation. Adjuvants are substances that can be used to augment a specific immune response. Normally, the adjuvant and the composition are mixed prior to presentation to the immune system, or presented separately, but into the same site of the mammal being elicited. Examples of materials suitable for use in vaccine compositions are provided in Oso, A., ed., Remington’s Pharmaceutical Sciences, Mack Publishing Co, Easton, Pa. (1980), pp. 1324-1341, which reference is entirely incorporated herein by reference.

[0981] Compositions comprising immunogenic minicells are injected into a human or animal at a dosage of 1-1000 µg per kg body weight. Antibody titers against growth factor are determined by ELISA, using the recombinant protein and horse-radish peroxidase-conjugated goat anti-human or animal immunoglobulins or other serologic techniques (e.g., sandwich ELISA). Booster injections are administered as needed to achieve the desired levels of protective antibodies and/or T cells.

[0982] Routes and frequency of administration, as well as dosage, will vary from individual to individual. Between 1 and 10 doses may be administered for a 52-week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. In immunotherapy of hyperproliferative disorders, a suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient’s tumor cells in vitro. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (e.g., more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients.

[0983] The vaccine according to the invention may contain a single species of immunogenic minicells according to the invention or a variety of immunogenic minicells, each of which displays a different immunogen. Additionally or alternatively, immunogenic minicells may each display and/or express more than one immunogen.

[0984] The summary of the invention described above is non-limiting and other features and advantages of the invention will be apparent from the following detailed description of the invention, and from the claims.

EXAMPLES

Example 1

Creation of a Minicell-Producing Bacterial Cell Line (MC-T7) that Expresses an Exogenous RNA Polymerase

[0985] In order to maximize the amount of RNA transcription from episomal elements in minicells, a minicell-producing cell line that expresses an RNA polymerase specific for certain episomal expression elements was created. This E. coli strain, designated MC-T7, was created as follows.

Recipient (P678-54) and donor (G43:BWS1619) strains were grown overnight in 10 mL of LB media (10 g NaCl, 10 g select peptone 140, and 5 g yeast extract in one liter ddH2O). The samples were centrifuged and then concentrated to about 0.2 mL of LB media. The concentrated samples were combined and incubated with slow rotation for 30 minutes at 30°C, and then were plated on LB agar plates that contained streptomycin (50 µg/mL) and tetracycline (50 µg/mL). (Ampicillin, streptomycin, tetracycline, and all other chemicals were purchased from Sigma Chemical (St. Louis, Mo.) unless otherwise indicated.) Recipient cells were resistant to streptomycin and donor cells were resistant to tetracycline; only conjugates, which contained both resistance genes, were able to grow on the LB agar plates that contained streptomycin (50 µg/mL) and tetracycline (50 µg/mL).

Putative conjugates were screened for Lambda phage sensitivity using a cross streak technique, in which putative colonies were cross-streaked on an LB agarose plate (streptomycin, 50 µg/mL, and tetracycline, 50 µg/mL) that had been streaked with live Lambda phage. The streaked conjugate colonies were streaked perpendicular to the Lambda phage streak; if a conjugate was sensitive to Lambda phage infection then, upon contact with the Lambda phage streak, there was cell lysis and thus less or no bacterial growth. Thus, in the case of conjugates that were sensitive to Lambda phage, there was deceased bacterial growth “down-streak” from the phage streak.

The conjugate E. coli that were found to be sensitive to Lambda phage infection were then used to create Lambda lysogens. Lysogenization is a process during which a Lambda phage incorporates its genome, include exogenous genes added thereto, into a specific site on the chromosome of its E. coli host cell.

The DE3 gene, which is present in the genome of Lambda phage used to create lysogens, encodes RNA polymerase from bacteriophage T7. Lysogenization was carried out using the DE3-Lysogenation kit (Novagen, Madison, Wis.) essentially according to the manufacturer's instructions. A T7 polymerase dependent tester phage was used to confirm the presence and expression of the DE3 gene on the bacterial chromosome. The T7-dependent tester phage can only form plaques on a bacterial strain in the presence of T7 polymerase. The phage uses a T7 promoter for expression of its essential genes. Therefore in a plaque-forming assay only cells which express T7 polymerase can be lysed by the tester phage and only these cells will allow for the formation of plaques. As is described in more detail herein, episomal expression elements that are used in minicells may be designed such that transcription and translation of a cloned gene is driven by T7 RNA polymerase by utilizing expression sequences specific for the T7 RNA polymerase.

Example 2

Cloning of Rat EDG-1 into the pCAL-C Expression Vector

Materials

Taq Polymerase, PCR Buffers, and PCR reagents were purchased from Roche Molecular Biochemicals (Indianapolis, Ind.). All restriction enzymes were purchased from Gibco BRL (Grand Island, N.Y.) and Stratagene (La Jolla, Calif.). QIAprep mini and maxi kits, PCR purification Kits, RNeasy miniprep kits, and the One Step RT-PCR Kit were purchased from QIAGEN (Valencia, Calif.). The Geneclean Kit was purchased from BIO 101 (Carlsbad, Calif.). IPTG (isopropyl-beta-D-thiogalactopyranoside), T4 DNA Ligase, LB Media components and agarose were purchased from Gibco BRL. The pCAL-c prokaryote expression vector and competent cells were purchased from Stratagene.

The pCAL-c expression vector has a structure in which an ORF may be operably linked to a high-level (but T7 RNA polymerase dependent) promoter, sequences that bind the E. coli Lac repressor, and the strong T7 gene 10 ribosome-binding site (RBS). The Lad repressor is also encoded by an expression from a pCAL-c vector. As long as it is bound to its recognition sequences in the pCAL-c expression element, the lac repressor blocks transcription from the T7 promoter. When an inducing agent, such as IPTG is added, the lac repressor is released from its binding sites and transcription proceeds from the T7 promoter, provided the T7 RNA polymerase is present. After induction, the cloned and expressed protein may constitute the majority of newly expressed cellular proteins due to the efficient transcription and translation processes of the system.

Amplification

The first step in cloning rat Edg-1 (rEDG-1) into an expression vector was to design primers for amplification via PCR (polymerase chain reaction). PCR primers were designed using the rat Edg-1 sequence (Nikajma et al., Biophy. J. 78:319 A, 2000) in such a manner that they contained either sites for NheI (GCATGC) or BamHII (GGATCC) on their five prime ends. The upstream primer had the sequence of SEQ ID NO:31. The three prime downstream primer (SEQ ID NO:32) also contained a stop codon, as the pCAL-c vector contains a Calmodulin Binding Protein (CBP) “tag” at its carboxyl terminus which was not intended to be incorporated into the rat Edg-1 polypeptide in this expression construct. The primer and resulting PCR products were designed so that the five prime end of the rat Edg-10RF was in frame with the methionine start codon found in the pCAL-c vector. Oligonucleotide Primer Sequences for Cloning into pCAL-C:

Edg1/pCAL-c construct primers:

Upstream primer

5'-AATTGCAAGCTTCCGCAACTGGAATCGGTCTTGAATTTGTA-3'

SEQ ID NO: 31

Downstream primer

5'-AATTGCAAGCTTGAAGAGGAATTGCAAGGT-3'

SEQ ID NO: 32
continued

EdgI/CFP fusion construct primers:
Upstream primer
(SEQ ID NO: 31)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

Downstream primer
(SEQ ID NO: 33)
5'-AATTGTACCGAGAACAGAAGGTAGTACCCCTTAACCA-3'

EdgI/Hin6 construct primers:
Upstream primer
(SEQ ID NO: 31)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

Downstream primer
(SEQ ID NO: 34)
5'-AATTGTACCGAGAACAGAAGGTAGTACCCCTTAACCA-3'

EdgI/rtPCR primers:
Upstream primer
(SEQ ID NO: 35)
5'-TTATGCAACACGACCGAAGGCAAGC-3'

Downstream primer
(SEQ ID NO: 36)
5'-AAGCCGTCACCCGCAAGGAC-3'

EdgI/pCAL-c construct primers:
Upstream primer
(SEQ ID NO: 37)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

Downstream primer
(SEQ ID NO: 38)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

EdgI/Hin6 construct primers:
Upstream primer
(SEQ ID NO: 39)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

Downstream primer
(SEQ ID NO: 40)
5'-AATTGTACTCCACCGATCCCGTTGATGTTAA-3'

GFP/pCAL-c construct primers:
Upstream primer
(SEQ ID NO: 41)
5'-GTTCCGCTACACGTTGGAACAGGCAA-3'

Downstream primer
(SEQ ID NO: 42)
5'-GTTCCGCTACACGTTGGAACAGGCAA-3'

GFP/CFP construct primers:
Upstream primer
(SEQ ID NO: 43)
5'-GTTCCGCTACACGTTGGAACAGGCAA-3'

Downstream primer
(SEQ ID NO: 43)
5'-GTTCCGCTACACGTTGGAACAGGCAA-3'

Notes:
Restriction endonuclease sites are underlined
Stop codons are double underlined

[0096] The primers were used to amplify the rEdg-1 DNA ORF using the polymerase chain reaction (PCR). The template used for amplification was mRNA isolated from rat muscle tissue using the RNeasy Miniprep Kit (Qiagen) and was carried out essentially according to the manufacturer’s protocol. Both the rPCR and PCR amplification steps were carried out in a single reaction using the OneStep RT-PCR Kit (Qiagen). The resulting rat Edg-1 PCR fragment was purified using the PCR Purification Kit (Qiagen). The amplified double stranded rEdg-1 DNA sequence contained the Nhel site at the 5’-prime end and the BamHI site at the 3’-prime end. This amplified rEdg-1 fragment was used for cloning into the pCAL-c expression vector.

[0097] The pCAL-c expression vector contains Ncol, Nhel, and BamHI restriction sites in its multiple cloning site. In order to insert rEdg-1 encoding sequence into the expression vector, the rEdg-1 PCR fragment and the pCAL-c expression vector were digested with Nhel and BamHI restriction enzymes for one hour at 37°C. The reaction mixture for the digestion step consisted of 1 µg of DNA, 1x restriction buffer, and 1 µl of each enzyme. The reaction mixture was brought to a final volume of 20 µl with ddH₂O (dd, double distilled). After 45 minutes, 1 µl of calf intestine alkaline phosphatase (CIAP) was added to the pCAL-c expression vector in order to remove the terminal phosphates from the digested plasmid DNA. The reactions were incubated for an additional 15 minutes at 37°C. The digested DNA samples were then run on a 1% TAE (Tris-acetate/EDTA electrophoresis buffer) agarose gel at 150 volts for 45 minutes. The bands were visualized with UV light after the gel was stained with ethidium bromide.

[0098] The appropriate bands were cut out of the gel for purification using the Geneclean Kit (BIO101). The purified DNA fragments were then quantified on a 1% TAE agarose gel. For the ligation reaction, ratios of insert to vector of 6:1 and 3:1 were used. A negative control comprising vector only was also included in the ligation reactions. The reactions consisted insert and vector DNA, 4 µl Ligase buffer, and 2 µl Ligase. The reaction was brought up to a final volume of 20 µl with ddH₂O. The ligation was carried out at room temperature for about 2 hours. Ten (10) µl of the ligation reaction mixture was used for subsequent transformation steps.

[0099] Ligated DNA was introduced into Epicurian Coli XL1-Blue competent cells using the heat shock transformation technique as follows. The ligation mixture was added to 100 µl of competent cells, placed on ice, and was incubated for about 30 minutes. The cells were then heat shocked at 37°C for 1 minute and put back on ice for 2 minutes. Following heat shock, 950 µl of room temperature LB media was added to the cells and the cells were shaken at 37°C for 1 hour. Following the 1-hour agitation the cells were pelleted for one minute at 12000 rpm in a Eppendorf 5417C microcentrifuge. The supernatant was carefully poured off so that about 200 µl remained. The cells were then resuspended in the remaining LB media and spread on 100x15 mm LB agarose plates containing 50 µg/ml ampicillin. The plates were incubated overnight at 37°C. Colonies were counted the following day, and the ratio of colonies between the negative control and the ligated samples was determined. A high ratio of the number of colonies when the ligation mixture was used to transform cells, as contrasted to the number of negative control colonies indicated that the ligation was successful. Transformed colonies were identified, isolated, and grown overnight in LB media in the presence of ampicillin. The resulting bacterial populations were screened for the presence of the Edg-1 pCAL-c expression construct.

[1000] Plasmid DNA was isolated from the cells using the QIAprep Spin Miniprep Kit (Qiagen). Isolated Edg-1 pCAL-c constructs were screened using the restriction
enzyme Apal, which digests the Edg-1-pCAL-c construct at two different sites: one in the Edg-1 coding sequence and one in the pCAL-c vector itself. The plasmid preparations were digested with Apal electrophoresed on a 1% TAE agarose gel and visualized using UV light and ethidium bromide staining. The predicted sizes of the expected DNA fragments were 2065 bp and 4913 bp. As shown in FIG. 3, bands of the predicted size were present on the gel. The entire Edg-1-pCAL-c construct was sequenced in order to confirm its structure. This expression construct, a pCAL-c derivative that contains the rat Edg-1 ORF operably linked to a T7 promoter and lac repressor binding sites, is designated “prEDG-1” herein.

Example 3

Construction of Rat EDG-1-CBP Fusion Protein

[1001] In order to detect rat Edg-1 protein expression, rEdg-1 coding sequences were cloned into the pCAL-c vector in frame with a CBP fusion tag. The cloning strategy for the rEdg-1-CBP construct was performed essentially as described for the Edg-1-pCAL-c construct with the following differences. The PCR primers (SEQ ID NO:35 and 36) were as described for the Edg-1-pCAL-c construct except for the omission of the stop codon in the downstream primer (SEQ ID NO:33). The removal of the stop codon is required for the construction of the Edg-1-CBP fusion protein. The pCAL-c vector is designed so that when the BamHI site is used for insertional cloning, and no stop codon is present in an ORF inserted into the pCAL-c expression vector the cloned ORF will be in-frame with the CBP fusion tag. Because the three prime downstream primer did not contain a stop codon, a CBP fusion tag could be cloned in-frame with the Edg-1 ORF. Other cloning steps were performed essentially as described before. The resulting plasmid, a pCAL-c derivative that comprises an ORF encoding a rat Edg-1-CBP fusion protein operably linked to a T7 promoter and lac repressor binding sites, is designated “prEDG-1-CBP” herein.

Example 4

Cloning of a His-Tagged Rat Edg-1 into pCAL-C Expression Vector

[1002] The rEdg-1 protein was manipulated to generate a fusion protein having a 6xHis tag at its carboxyl terminus. A “6xHis tag” or “His tag” is an amino acid sequence consisting of six contiguous histidine residues that can be used as an epitope for the binding of anti-6xHis antibodies, or as ligand for binding nickel atoms. The His-tagged rEdg-1 fusion protein is used to detect rEdg-1 protein expression in the minicell expression system environment.

[1003] The Edg-1-6xHis construct was cloned using the strategy described above for the construction of the rEdg-1-pCAL-c expression construct (prEDG-1), with the upstream primer having the sequence of SEQ ID NO:3, but with the exception that the three prime downstream primer (SEQ ID NO:34) was designed to contain six histidine codons followed by a stop codon. The 18 base pair 6xHis tag was incorporated into the carboxyl terminus of the Edg-1 protein as expressed from the pCAL-c vector. Subsequent cloning procedures (PCR, restriction digest, gel purification, ligation, transformation, etc.) were performed as described previously for the Edg-1-pCAL-c construct (prEDG-1). The resulting plasmid, a pCAL-c derivative that comprises an ORF encoding a carboxy-terminal His-tagged rat Edg-1-CBP fusion protein operably linked to a T7 promoter and lac repressor binding sites, is designated “prEDG-1-6xHis” herein.

Example 5

Amplification and Cloning of Rat EDG-3 Sequences

[1004] The Edg-3 full length coding sequence was amplified via PCR from rat skeletal muscle mRNA using primers (SEQ ID NO:35 and 36) designed from the known mouse sequence (Genbank accession NM_010101). The mRNA was used as a template for the amplification reaction which was isolated using the RNeasy Miniprep Kit (Qiagen). Both the PCR and PCR amplification steps were carried out in a single reaction using the One Step RT-PCR Kit (Qiagen). The rEdg-3 PCR products were visualized with UV after electrophoresis in 1% TAE agarose gels and ethidium bromide staining.

[1005] The predicted size of the amplified PCR products is 1145 base pairs. An appropriately-sized DNA band was isolated from the TAE gel and purified using the GeneClean Kit (B10101). The purified band was ligated to the pCR3.1 vector using the TA-cloning kit (Invitrogen). Other cloning steps were carried out as described previously for the cloning of the rEdg-1-pCAL-c construct (prEDG-1) with the exception that the samples were screened using the EcoRI restriction enzyme. The expected sizes of the digested bands were 1145 base pairs and 5060 base pairs. Positive clones were analyzed by automated sequencing. The nucleotide sequences were analyzed using BLAST searches from the NCBI web site (www.ncbi.nlm.nih.gov/). The predicted full length rat Edg-3 amino acid sequence was assembled from the nucleotide sequencing data using in silico translation. The pCR3.1 vector comprising the rat Edg-3 ORF is designated “pCR-rEDG-3” herein.

Example 6

Cloning of Rat EDG-3 Coding Sequences into the pCAL-C Expression Vector

[1006] In order to express it in the minicell expression system, the rat Edg-3 ORF was cloned into the pCAL-c expression vector. The cloning strategy used was as described above for the cloning of the rat Edg-1 gene into the pCAL-c vector with the following exceptions. The primers used for PCR amplification were designed from the rat Edg-3 sequence and contained sites for the restriction enzymes Nhel and Kpn1 (GGTACC). The Nhel site was added to the five prime upstream primer (SEQ ID NO:37) and the Kpn1 site was added to the three prime downstream primer; SEQ ID NO:38). The Nhel and Kpn1 restriction enzymes were used for the digestion reaction. The reaction mixture for the digestion step consisted of 1 μg of DNA, 1x restriction buffer (provided with the enzyme), and 1 μL of each enzyme. Plasmid preparations were screened by digestion with Nhel and Kpn1. The digested plasmid DNA was electrophoresed on a TAE agarose gel and visualized by UV after staining with ethidium bromide. The resultant band sizes were predicted to be 1145 base pairs and 5782 base pairs. The positive plasmid clones were analyzed with automated sequencing. The resulting plasmid, a pCAL-c derivative that comprises an ORF
encoding a rat Edg-3 protein operably linked to a T7 promoter and lac repressor binding sites, is designated “pEDG-3” herein.

Example 7
Cloning of a His-Tagged Rat Edg-3 into the pCAL-C Expression Vector

In order to detect expression of the rat Edg-3 protein in the minicell expression system, the rat Edg-3 coding sequence was manipulated so as to contain a 6xHis tag at the carboxyl terminus of the protein. The cloning strategy used to create this construct was essentially the same as described above for the rEdg-3-pCAL-c (pEDG-3) construct cloning, with the upstream primer having the sequence of SEQ ID NO:37, with the exception that the three-prime downstream primer (SEQ ID NO:18) was designed to contain a 6xHis coding sequence followed by a stop codon, which allowed for the incorporation of the 6xHis amino acid sequence onto the carboxyl terminus of the Edg-3 receptor protein. Other cloning and screening steps were performed as described above. The resulting plasmid, a pCAL-c derivative that comprises an ORF encoding a carboxy-terminal His-tagged rat Edg-3 fusion protein operably linked to a T7 promoter and lac repressor binding sites, is designated “pEDG-3-6xHis” herein.

Example 8
GFP Cloning into pCAL-C Expression Construct

Cloning of GFP-encoding nucleotide sequences into the pCAL-c vector was performed in order to produce an expression construct having a reporter gene that can be used to detect protein expression (GFP, green florescent protein). The cloning strategy used was essentially the same as the cloning strategy described above with the following exceptions. The template used for PCR amplification was the pCAGGS plasmid “construct” (GFP construct sold by Clontech). The primers used for amplification were designed from the GFP coding sequence and contained sites for the restriction enzymes Neol and BamHI. The Neol site was added to the five prime upstream primer (SEQ ID NO:40) and the BamHI site was added to the three prime downstream primer, see SEQ ID NO:41) The Neol and BamHI restriction enzymes were used for the digestion reaction. The reaction mixture for the digestion step consisted of 1 µg of DNA, 1× restriction buffer (provided with the enzyme), and 1 µL of each enzyme. The screening of the plasmid preparations was carried out using Neol and BamHI. Digested plasmid preparations were electrophoresed and visualized on TAE agarose gels with UV after staining with ethidium bromide. Restriction products having the predicted sizes of 797 and 5782 base pairs were seen. Positive plasmid clones were sequenced using an automated sequencer. The resulting plasmid, a pCAL-c derivative that comprises an ORF encoding a rEdg-3-GFP fusion protein operably linked to a T7 promoter and lac repressor binding sites, is designated “pEDG-3-GFP” herein.

Example 9
Design Construction of Control Expression Elements

Control expression elements used to detect and quantify expression of proteins in minicells were prepared. These controls direct the expression of detectable proteins. An expression element used as positive control is pPTC12, which is supplied with the pC AL-c expression vector from Stratagene. This construct contains an ORF encoding a fusion protein comprising beta-galactosidase linked to CBP. Induction of expression of pPTC12 should result in the production of a protein of about 120 kD, and this protein is detected via its enzymatic activity or by using antibodies directed to epitopes on the beta-galactosidase or CBP polypeptide.

A GFP fusion construct was created and used as a positive control for the CBP detection kit. This construct was a positive control for induction of protein expression in the minicell expression system. The cloning strategy used to create the construct was essentially the same as that used for the cloning of the GFP into the pCAL-c expression vector, with the exception that the three prime downstream primer did not contain a stop codon; this allowed for the in frame incorporation of the CBP fusion tag to the GFP protein. The upstream primer had the sequence of SEQ ID NO:42, and the downstream primer had the sequence SEQ ID NO:43. The nucleotide sequence of the expression element was confirmed using an automated sequencer. The resulting plasmid, a pCAL-c derivative that comprises an ORF encoding GFP operably linked to a T7 promoter and lac repressor binding sites, is designated “pGFP-CBP” herein.

Example 10
Introduction of pCAL-C Expression Constructs into the MC-T7 Escherichia coli Strain

The MC-T7 E. coli strain was made competent using the CaCl₂ technique. In brief, cells were grown in 40 mL LB medium to an OD600 of 0.6 to 0.8, and then centrifuged at 8000 rpm (7,700 g) for 5 min at 4°C. The pellet was resuspended in 20 mL of cold CaCl₂ and left on ice for five minutes. The cells were then centrifuged at 8000 rpm (7,700 g) for 5 min at 4°C. The cell pellet was resuspended in 1 mL of cold CaCl₂ and incubated on ice for 30 min. Following this incubation 1 mL of 25% glycerol was added to the cells and they were distributed and frozen in 200 µL aliquots. Liquid nitrogen was used to freeze the cells. These cells subsequently then used for the transformation of expression constructs.

Example 11
Preparation of Minicells

To some degree, the preparation of minicells varied according to the type of expression approach that is used. In general, there are two such approaches, although it should be noted from the outset that these approaches are neither limiting nor mutually exclusive. One approach is designed to isolate minicells that already contain an expressed therapeutically protein or nucleic acid. Another approach is designed to isolate minicells that will express the protein or nucleic acid in the minicell following isolation.

E. coli are inoculated into bacterial growth media (e.g., Luria broth) and grown overnight. After this, the overall protocol varies with regards to methods of induction of expression. The minicell producing cultures used to express protein post isolation are diluted and grown to the desired OD600 or OD450, typically in the log growth phase of bacterial cultures. The cultures are then induced with IPTG and then isolated. The IPTG concentration and exposure...
depended on which construct was being used, but was usually about 500 μM final for a short time, typically about 4 hours. This treatment results in the production of the T7 polymerase, which is under control of the LacUV5 promoter, which is repressed by the Lac repressor protein. IPTG relieves the Lac repression and thus induces expression from the LacUV5 promoter which controls expression of the T7 polymerase from the chromosome. This promoter is “leaky” that is, there is always a basal level of T7 polymerase which can be selected for or against so that the induction before isolation is not required. (This induction step is not required if a non-T7 expression system is used, as the reason for this step is to express the T7 RNA polymerase in the minicell-producing cells so that the polymerase and molecules segregate with the minicell.)

[1014] The E. coli cultures that produce minicells containing a therapeutic protein or nucleic acid have different induction protocols. The overnight cultures are diluted as described above; however, in the case of proteins that are not toxic to the parent cells, this time the media used for dilution already contains IPTG. The cultures are then grown to mid-log growth and minicells are isolated. These cultures produce the therapeutic protein or nucleic acid as they grow, and the minicells derived therefrom contain the therapeutic protein or nucleic acid.

[1015] Alternatively or additionally, IPTG is added and expression is induced after the isolation of minicells. In the case of non-toxic proteins or nucleic acids that are expressed from expression elements in minicells, this treatment enhances production of the episomal encoded gene product. In the case of toxic gene products induction post-isolation is preferred.

Example 12
Minicell Isolation

[1016] Minicells were isolated from the minicell producing MC-T7 strain of E. coli using centrifugation techniques. The protocol that was used is essentially that of Janatiapour et al. (Translocation of Vibrio Harveyi NN-N-Diacytetylcholase to the Outer Membrane of Escherichia coli, J. Bacteriol. 169: 3785-3791, 1987) and Matsunuma et al. (Synthesis of Mot and Che Products of Escherichia coli Programmed by Hybrid ColE1 Plasmids in Minicells, J. Bacteriol. 132:996-1002, 1977).

[1017] In brief, MC-T7 cells were grown overnight at 37°C in 2 to 3 ml of LB media containing ampicillin (50 μg/mL), streptomycin (50 μg/mL), and tetracycline (50 μg/mL) (ampicillin was used only when growing MC-T7 cells containing a pCAL-c expression construct). The cells were diluted 1:100 in a total volume of 100 to 200 ml LB media with antibiotics, and grown at 37°C until they reached an OD₆₀₀ of 0.4 to 0.6, which is roughly beginning of the log growth phase for the MC-T7 E. coli. During this incubation the remainder of the overnight culture was screened for the presence of the correct expression construct using the techniques described above. When the cultures reached the appropriate OD₆₀₀, they were transferred to 250 ml GS3 centrifuge bottles and centrifuged (Beckman centrifuge) at 4500 rpm (3,500 g) for 5 min. At this point the supernatant contains mostly minicells, although a few relatively small whole cells may be present.

[1018] The supernatant was transferred to a clean 250 ml GS3 centrifuge bottle and centrifuged at 8000 rpm (11,300 g) for 10 min. The pellet was resuspended in 2 ml of 1xBSG (10xBSG: 85 g NaCl, 3 g KH₂PO₄, 6 g Na₂HPO₄, and 1 g gelatin in 1 L ddH₂O) and layered onto a 32 ml 5 to 20% continuous sucrose gradient. The sucrose gradient was made with sucrose dissolved in 1xBSG.

[1019] The sucrose gradient was then loaded in a Beckman SW24 rotor and centrifuged in a Beckman Ultracentrifuge at 4500 rpm (9,000 g) for 14 min. Following ultracentrifugation a single diffuse band of minicells was present. The top two thirds of this band was aspirated using a 10 ml pipette and transferred to a 30 ml Oakridge tube containing 10 ml of 1xBSG. The sample was then centrifuged at 13,000 rpm (20,400 g) for 8 min. Following centrifugation, the pellet was resuspended in 2 ml 1xBSG, and the resuspended cells were loaded onto another 5 to 20% sucrose gradient. This sucrose gradient was centrifuged and the minicells were collected as described above. The sucrose gradient procedure was repeated a total of three times.

[1020] Following the final sucrose gradient step the entire minicell band was collected from the sucrose gradient and added to a 30 ml Oakridge tube that contained 10 ml of MMM buffer (200 ml 1xM2 salts, 2 ml 20% glucose, and 2.4 ml DIFCO Methionine Assay Medium). This minicell solution was centrifuged at 13,000 rpm (20,400 g) for 8 min. The pellet was resuspended in 1 ml of MMM Buffer.

[1021] The concentration of minicells was determined using a spectrophotometer. The OD₂₆₀ was obtained by reading a sample of minicells that was diluted 1:100.

Example 13
Other Methods to Prepare and Isolate Minicells

[1022] By way of non-limiting example, induction of E. coli parental cells to form minicells may occur by overexpression of the E. coli ftsZ gene. To accomplish this both plasmid-based and chromosomal overexpression constructs were created that place the ftsZ gene under the control of various regulatory elements (Table 6).

<table>
<thead>
<tr>
<th>Regulatory region</th>
<th>inducer</th>
<th>[inducer]</th>
<th>SEQ ID NO.:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pm:n:ftsZ</td>
<td>Arabinose</td>
<td>10 nM</td>
<td>1,3</td>
</tr>
<tr>
<td>Pfn:ftsZ</td>
<td>Rhamnose</td>
<td>1 nM</td>
<td>2,4</td>
</tr>
<tr>
<td>Ptn:ftsZ</td>
<td>IPTG</td>
<td>30 μM</td>
<td>5, Garrido et al.*</td>
</tr>
</tbody>
</table>


Oligonucleotide Names and PCR Reactions Use the Following Format:

[1023] “gene-1” is N-terminal, 100% homology oligo for chromosomal or cDNA amplification

[1024] “gene-2” is C-terminal, 100% homology oligo for chromosomal or cDNA amplification

[1025] “gene-1 RE site” is same sequence as gene-1 with additional residues for remainder of sequence, RE sites, and/or chimeric fusions.

[1026] “gene-2-RE site” is same sequence as gene-1 with additional residues for remainder of sequence, RE sites, and/or chimeric fusions.
Use “gene-1, 2” combo for chromosomal/cDNA amplification and “gene-1 RE site, gene-2-RE site” to amplify the mature sequence from the “gene-1, 2” gel-purified product.

<table>
<thead>
<tr>
<th>SEQ</th>
<th>Primer NO.: name 5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>FtsZ-1 CCAATGGAACCTACCAATGACCCG</td>
</tr>
<tr>
<td>45</td>
<td>FtsZ-2 GCCCTGTACAGGAAATGCTGGG</td>
</tr>
<tr>
<td>46</td>
<td>FtsZ-2-1 CGGCGCTGCAAGATTTGTTAAGCCGACCTGACCACCTG</td>
</tr>
<tr>
<td></td>
<td>PstI AAAGAGCGC</td>
</tr>
<tr>
<td>47</td>
<td>FtsZ-2-2 GGCCTCTGAGATTAATACGCGTTCCTAGCCG</td>
</tr>
<tr>
<td></td>
<td>XbaI GAAAGCTGG</td>
</tr>
</tbody>
</table>

Table 7 Oligonucleotide Sequences are for Use in Cloning ftsZ into SEQ ID NO: 1 and 2 (Insertions of ftsZ Behind the Arabinose Promoter (SEQ ID NO.: 1) and the Rhamnose Promoter (SEQ ID NO.: 2))

<table>
<thead>
<tr>
<th>SEQ</th>
<th>Primer NO.: name 5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>Kan-1 GCCGAACTGGGCGGGTTTTATGGAAGCAAGC</td>
</tr>
<tr>
<td>49</td>
<td>Kan-2 GCCGAACTGGGAGAAGAGCTGCCAGGAGCC</td>
</tr>
<tr>
<td>50</td>
<td>Kan-1 GGGCGCTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td></td>
<td>X-frag AGCTCTATATCCTAGAAGGGAATGCG</td>
</tr>
<tr>
<td></td>
<td>TCTAGGAGAAGGATAGAAG</td>
</tr>
<tr>
<td>51</td>
<td>Kan-2-1 GAAAGCTGGGCGGGTTTTATGGAAGCAAGC</td>
</tr>
<tr>
<td></td>
<td>X-frag GGGCGCTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td>52</td>
<td>AraC-1 GGGCGCTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td>53</td>
<td>RhaR-1 TTACAGGAAATGGAAGC</td>
</tr>
<tr>
<td>54</td>
<td>LacI-1 GTGAGCTCATATTTT</td>
</tr>
<tr>
<td></td>
<td>CAGGGAAGAGAAAGAAG</td>
</tr>
<tr>
<td>55</td>
<td>LacI-1-1 GTGAGCTCATATTTT</td>
</tr>
<tr>
<td></td>
<td>CAGGGAAGAGAAAGAAG</td>
</tr>
<tr>
<td>56</td>
<td>AraI-1 GCCGCGTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td></td>
<td>TCCGTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td>57</td>
<td>RhaR-1 GCCGCGTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td></td>
<td>TCCGTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td>58</td>
<td>PtsZ-1 GCCGCGTTACAGGAATGCTGGG</td>
</tr>
<tr>
<td></td>
<td>TCCGTTACAGGAATGCTGGG</td>
</tr>
</tbody>
</table>

[1028] In like fashion, the ftsZ gene was amplified from SEQU ID NO.: 1, 2 and Ptc::ftsZ (Garrido, T. et al. 1993. Transcription of ftsZ oscillates during the cell cycle of Escherichia coli. EMBO J. 12:3957-3965) plasmid and chromosomal constructs, respectively using the following oligonucleotides:

For amplification of araC through ftsZ of SEQ ID NO.: 1 use oligonucleotides:

AraC-1
FtsZ-2

[1029] For amplification of rhaR through ftsZ of SEQ ID NO.: 2 use oligonucleotides:

RhaR-1
FtsZ-2

[1030] For amplification of lacI through ftsZ of Ptc::ftsZ (Garrido, T., et al.) use oligonucleotides:

lacI-1
ftsZ-2

[1031] The above amplified DNA regions were gel-purified and used as template for the second round of PCR using oligonucleotides containing homology with the E. coli chromosomal gene intD and on the other end with random sequence termed “X”. Oligonucleotides used in this round of PCR are shown below:

For amplification of araC through ftsZ from SEQ ID NO.: 1 to contain homology to intD and the random X use oligonucleotides:

AraC-1-intD
FtsZ-1-X

[1032] For amplification of rhaR through ftsZ from SEQ ID NO.: 2 to contain homology to intD and the random X use oligonucleotides:

RhaR-1-intD
FtsZ-1-X

[1033] For amplification of lacI through ftsZ from Ptc::ftsZ to contain homology to intD and the random X use oligonucleotides:

lacI-1-intD
FtsZ-1-X

[1034] The PCR products from these PCR reactions are as shown below:

[1035] intD-araC-Ara promoter-ftsZ—“X”
[1036] intD-rhaR-Rha promoter-ftsZ—“X”
[1037] intD-lacI-Ptc promoter-ftsZ—“X”

To amplify the mature complexes, the following regions were mixed and amplified with the coupled oligonucleotide sequence primers:
These expression constructs may be expressed from the plasmid, placed in single copy, replacing the native ftsZ copy on the E. coli chromosome (Garrido, T., et al. 1993. Transcription of ftsZ oscillates during the cell cycle of Escherichia coli. EMBO J. 12:3957-3965), or in duplicate copy retaining the native ftsZ copy while inserting one of the expression constructs in Table 6 into the intD gene on the same chromosome. Chromosomal duplications were constructed using the RED recombinase system (Katsenko, K. A., and B. L. Wanner. One-Step Inactivation of Chromosomal Genes in Escherichia coli K-12 Using PCR Products. Proc. Natl. Acad. Sci. 97:6640-6645, 2000) and are shown in SEQ ID NO 3-5. The later constructs allow native replication during non-minicell producing conditions, thus avoiding selective pressure during strain construction and maintenance. Furthermore, these strains provide defined points of minicell induction that improve minicell purification while creating conditions that allow strain manipulation prior to, during, and following minicell production. By way of non-limiting example these manipulations may be protein production that the cytoplasmic redox state, modify plasmid copy number, and/or produce chaperone proteins.

For minicell production, a minicell producing strain described in the previous section is grown overnight in Luria broth (LB) supplemented with 0.1% dextrose, 100 μg/ml ampicillin, and when using the single-copy ftsZ construct, 15 μM IPTG. All incubations were performed at 37° C. For minicell induction only, overnight strains are subcultured 1/1000 into the same media. If minicell induction is to be coupled with co-expression of other proteins that are controlled by a catabolite repression-sensitive regulator, dextrose was avoided. Minicell induction is sensitive to aeration and mechanical forces. Therefore, flask size, media volume and shake speed is critical for optimal yields. Likewise, bioreactor conditions must be properly regulated to optimize these production conditions.

In shake-flask cultures, strains are grown to early exponential (log) phase as monitored by optical density (OD) at 600 nm (OD600 0.05-0.20). (Bioreactor conditions may differ significantly depending on the application and yield desired). For minicell induction alone, early log phase cultures are induced with the appropriate inducer concentration shown in Table 6. For coupled co-expression, these cultures are induced as shown in Table 6 for the appropriate minicell regulator, while the coupled protein(s) is induced with the inducer appropriate for the regulator controlling the synthesis of that protein. Cultures are grown under the appropriate conditions and harvested during late log (OD600 0.8-1.2). Depending on the application, minicell induced cultures may be immediately chilled on ice prior to purification, or maintained at room temperature during the harvesting process.

To separate minicells from viable, parental cells, cultures are subjected to differential centrifugation (Voros, J., and R. N. Goodman. 1965. Filamentous forms of Erwinia amylovora. Phytopathol. 55:876-879). Briefly, cultures are centrifuged at 4,500 rpm in a GSA rotor for 5 min. Supernatants are removed to a fresh bottle and centrifuged at 8,000 rpm for an additional 10 min to pellet minicells. Pelleted minicells (containing contaminating parental cells) are resuspended in 2 ml L.B. (LB supplemented with 0.1% dextrose), Min (minimal M63 salt media) (Roozen, K. J., et al. 1971. Synthesis of ribonucleic acid and protein in plasmid-containing minicells of Escherichia coli K-12. J. Bacteriol. 107:21-23), supplemented with 0.5% casamino acids) or MDT (minimal M63 salt media, supplemented with 0.5% casamino acids, 0.1% dextrose, and thiamine). Resuspended
minicells are next separated using linear density gradients. By way of non-limiting example, these gradients may contain sucrose (Cohen A., et al. 1968. The properties of DNA transferred to minicells during conjugation. Cold Spring Harb. Symp. Quant. Biol. 33:635-641). ficoll, or glycerol. For example, linear sucrose gradients range from 5-20% and are poured in LB, LBD, Minor MDT. Using a SW28 swinging bucket rotor, gradients are centrifuged at 4,500 rpm for 14 min. Banded minicells are removed, mixed with LBD, LBD, Minor MDT; and using a JA-20 rotor are centrifuged at 13,000 rpm for 12 min. Following centrifugation, pellets are resuspended in 2 ml LB, LBD, Minor MDT and subjected to a second density gradient. Following the second density separation, banded minicells are removed from the gradient, pelleted as described, and resuspended in LB, LBD, Minor MDT for use and/or storage.

[1042] Purified minicells are quantitated using an OD_{600} measurement as compared to a standard curve incorporating LPS quantity, minicell size, and minicell volume. Quantitated minicells mixtures are analyzed for contaminating, viable parental cells by plating on the appropriate growth media (Table 9).

<table>
<thead>
<tr>
<th>Purification</th>
<th>Total cells</th>
<th>Total parental cells</th>
<th>MC/PC ratio</th>
<th>Fold-purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>4.76 x 10^{11}</td>
<td>5.14 x 10^{10}</td>
<td>0.25%</td>
<td>—</td>
</tr>
<tr>
<td>After</td>
<td>1.49 x 10^{11}</td>
<td>6.01 x 10^{10}</td>
<td>2.48 x 10^{0}/1</td>
<td>5.23 x 10^{0}</td>
</tr>
</tbody>
</table>

Example 14
Protoplast Formation

[1043] In order to allow a membrane receptor to be presented to the outside environment (displayed), minicells are made into protoplasts. In order to make the integral membrane protein receptors in the inner membrane more accessible for ligand binding, the outer membrane and cell wall were removed. The removal of the outer membrane and cell wall from E. coli whole cells and minicells to produce protoplasts was performed essentially according to previously described protocols with a few modifications (Birdsell et al., Production and Ultrastructure of Lysozyme and Ethylenediaminetetraacetate-Lysozyme Spheroplasts of Escherichia coli, J. Bacteriol. 93:427-437, 1967; Weiss et al., Protoplast Formation in Escherichia Coli, J. Bacteriol. 128:668-670, 1976). Both minicells and whole cells were processed the same way.

In brief, the cells were grown to mid-log phase and pelleted at room temperature (minicells were isolated from cultures in mid-log phase). The pellet was washed twice with 10 mM Tris. Following the second wash protoplast production may be performed using two approaches. In the first approach, following the second wash, the cells were resuspended in 100 mM Tris (pH 8.0) that contained 6-20% sucrose and put in a 37°C waterbath (the Tris/sucrose buffer was pre-warmed to 37°C). The volume used to resuspend the cells was determined by the following equation: (volume of cells xOD_{600})/10-resuspension volume. After a 1 minute incubation, 2 mg/ml lysozyme was added to a final concentration of 5-100 μg/mL. The samples were then incubated for 12 minutes at 37°C. while being gently mixed. Next, 100 mM EDTA (pH 7) was slowly added over a period of 2.5 minutes (amount of EDTA added=1/100-1/10 volume of cells) followed by a 10 min incubation at 37°C. The protoplasts are also diluted from 20% sucrose down to either 10% or 5% sucrose, which facilitates the complete removal of the outer membrane and cell wall. The protoplasts thus generated were separated from the outer membrane and cell wall using a sucrose step gradient. A sucrose step gradient does not have a gradual increase in sucrose percentage; rather, it goes directly from one percent to the other. For example, protoplasts generated from whole cells are loaded on a step gradient that is made from 5% and 15% sucrose. The protoplasts spin through the 15% sucrose but the debris generated when making the protoplasts does not spin through the 15% sucrose. The protoplasts are thus separated from the debris. The second method to prepare protoplasts, following the second wash, 1x10^{10} cells were resuspended with 50 mM Tris, pH 8.0 containing 0.5-50 mM EDTA and 6-20% sucrose. This mixture was incubated at 37°C for 10 min. Following incubation, the mixture was centrifuged at 13,200 RPM in a microcentrifuge for 2 min. After centrifugation, the pellet was resuspended in 50 mM Tris, pH 8.0 containing 5-100 μg/ml lysozyme and 6-20% sucrose. This mixture was incubated at 37°C for 10 min. Following incubation, the mixture was centrifuged at 13,200 RPM in a microcentrifuge for 2 min, resuspended in 50 mM Tris pH 8.0 containing 6-20% sucrose for use.

[1044] An alternative method to remove contaminating LPS is to use affinity absorption with an anti-LPS antibody (Cortex). To accomplish this, the anti-LPS antibody was coated on either an activated agarose or sepharose matrix (Sigma) or epoxy-coated magnetic M-450 beads (Dynal). The spheroplast/protoplast mixture was subjected to the antibody coated matrix either in batch or using column chromatographic techniques to remove contaminating LPS. Following exposure, the unbound fraction(s) was collected and re-exposed to fresh matrix. To monitor the efficiency of the protoplasting reaction and LPS removal, three constructs were used (Table 10).

<table>
<thead>
<tr>
<th>PROTOPLAST MONITORING CONSTRUCTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
</tr>
<tr>
<td>PMPX-5</td>
</tr>
<tr>
<td>PMPX-5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Oligonucleotide Primer Sequences for Table 10 Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>59</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>61</td>
</tr>
<tr>
<td>PeT1</td>
</tr>
</tbody>
</table>

May 31, 2012
Oligonucleotides SEQ ID Nos.:59, 60, 61 and 62 were used to amplify phoA lacking a leader sequence (ΔphoA) form the E. coli chromosome. Once amplified, this region was inserted into SEQ ID NO.: 6 using PstI and XbaI to create SEQ ID NO.: 7.

Oligonucleotides SEQ ID Nos.:63, 64, 65 and 66 were used to amplify phoA containing a leader sequence (phoA) form the E. coli chromosome. Once amplified, this region was inserted into SEQ ID NO.: 6 using PstI and XbaI to create SEQ ID NO.: 8.

Oligonucleotides SEQ ID Nos.:59, 60, 67 and 68 were used to amplify phoA lacking a leader sequence (ΔphoA) form the E. coli chromosome and form a translational fusion between the transmembrane domain of toxR from Vibrio cholerae. Once amplified, this region was inserted into SEQ ID NO.: 6 using PstI and XbaI to create SEQ ID NO.: 9.

By co-expression of minicells and protein, minicells were prepared that contained cytoplasmic PhoA (pMPX-32 expresses phoA lacking a leader sequence [ΔphoA], periplasmic PhoA (pMPX-53 expresses native phoA that exports to the periplasmic space), or inner membrane-bound PhoA (pMPX-33 expresses phoA lacking a leader sequence fused to the transmembrane domain (TMD) of the toxR gene product from Vibrio cholerae). Using these expressed proteins, the efficiency of minicell protoplasting was monitored (Table 12).

Efficiency of Minicell Protoplast Preparation and Purification

<table>
<thead>
<tr>
<th>Step</th>
<th>Location</th>
<th>ΔPhoA</th>
<th>PhoA</th>
<th>Δ-ToxR PhoA</th>
<th>LPS total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minicell</td>
<td>Pellet</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>EDTA/lysozyme</td>
<td>Whole</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Efficiency of Minicell Protoplast Preparation and Purification

Example 15

T7-Dependent Induction of Expression

Expression from the pCAL-c expression vector is driven from a T7 bacteriophage promoter that is repressed by the Lac gene product. Transcription of the DNA into mRNA, and subsequent translation of mRNA into proteins, does not occur as long as the Lac repressor is bound to the T7 promoter. However, in the presence of IPTG, the Lac repressor does not bind the T7 promoter. Thus, induction of expression from pCAL-c sequences is dependent on the presence of IPTG. Slightly different protocols were used for the induction of Escherichia coli whole and for the induction of minicells. Slight differences are also present in the protocols for induction of minicells for 35S-methionine labeling of proteins in contrast to those for the induction of minicells for Western blot analysis. These induction protocols are described below.

For expression in E. coli whole cells, the cells were first grown overnight in 3 mL of LB and antibiotics. The cultures were screened for the presence of the desired expression element as previously described. Cultures containing the desired expression elements were diluted 1:100 and grown to an OD600 of between 0.4 to 0.6. The culture size varied depending on the intended use of the cells. IPTG was then added to a final concentration of 200 μg/mL, and the cells were shaken at 30°C for 4 hours. Following the induction, cells were harvested for analysis.

The induction of minicells was carried out as follows. The minicells were diluted in MMB buffer to 1 mL total volume according to the concentration obtained from the isolation procedure (OD450 of about 0.5). The cells were then treated with 50 μg/mL of cycloserine for 30 minutes at 37°C to stop whole cell growth. Following the cycloserine treatment the cells were provided with an amino acid, methionine, which the MMB buffer does not contain. For 35S-labeled protein induction 35S-methionine was added to the minicell sample whereas, for unlabeled protein induction unlabeled methionine was added. Fifteen (15) μCi of 35S-methionine (Amersham Pharmacia Biotech, Piscataway, N.J.) was added to the samples for radiolabeling and 5 μmol of methionine was added to the non-labeled minicell samples. Two hundred (200) μg/mL IPTG was also added to the minicell samples,
which were then shaken at 30°C for about 4 hours. Following induction, the minicells were harvested for further preparation or analysis.

Example 16

Western Blot Analysis

The CBP detection kit was purchased from Stratagene. SDS running buffer, 10% Tris-HCl ready gels, Kaleidoscope Pre-stained Standards, and Laemmi Sample Buffer were purchased from BIO RAD ( Hercules, Calif.). GFP (FL) HRP antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, Calif.). Edg-3CT antibody an antibody directed to the carboxy terminus of was purchased from Exolpia Biologicals (Boston, Mass.). Anti-6xHis antibody, postiso, and the WesternBreeze Kit were purchased from Invitrogen (Carlsbad, Calif.). Protocols were carried out essentially according to the manufacturer's instructions unless otherwise indicated.

Three different Western blot protocols were used to detect protein expression in both a minicell expression system and in a whole cell expression system. For both systems, the SDS-PAGE gel and the transfer protocols were essentially as follows. The samples were denatured by diluting the samples 1:1 in Laemmi buffer (BIORAD) and then sonicated for 10 min. The denatured samples were loaded onto a 10% Tris-Glycine gel ( BIORAD) and electrophoresed at 130 V for about 1.5 hours in SDS running buffer ( BIORAD). The electrophoresed proteins were electrotransferred to nitrocellulose membranes at 0.5 Amps for 1.5 hours in Transfer Buffer (5.8 g Tris, 2.9 g glycine, 200 mL methanol, and 3.7 mL of 10% SDS). The nitrocellulose membranes comprising the transferred proteins were carried out as follows. The nitrocellulose membrane was blocked for 2 hours with 5% milk in PBST (PBS buffer with 0.05% Tween). Following the blocking step the nitrocellulose membrane was washed twice with PBST. For the detection of GFP protein, an anti-GFP- HRP conjugated antibody ( Santa Cruz Biotechnology) was used at a dilution of 1:5000 in PBST (HRP, horse radish peroxidase). The nitrocellulose membrane was incubated in the anti-GFP-HRP antibody solution for one hour and then washed twice with PBST. GFP proteins on the nitrocellulose membrane were detected and visualized using the ECL system (Amersham).

The His-tagged Edg-1 and Edg-3 proteins were detected using a mouse anti-6xHis antibody from Invitrogen and the WesternBreeze chemolumincent Kit (Invitrogen). The antibody was diluted 1:4000 in buffers provided by the WesternBreeze Kit. The WesternBreeze immunoblot was carried out essentially according to the manufacturer’s protocol. The Edg-1CBP and GFP-CBP fusion proteins were detected using the CBP detection Kit (Stratagene). All antibodies and substrates were provided in the Kit. FIG. 3 is a photo of the Western hybridization results showing the presence of Edg-1-6xHis and Edg-3-6xHis in minicells and parent cells.

Example 17

Methods to Induce Expression

Expression in minicells may proceed following purification of minicells and/or minicell protoplasts from parental cells and LPS constituents, respectively. However, for some applications it is suitable to co-express proteins of interest with minicell induction. For these approaches, one may use the protocol described in EXAMPLE 13 for expression of the phoA constructs. By way of non-limiting example, either of these approaches may be accomplished using one or more of the following expression constructs (Table 13).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Regulatory element(s)</th>
<th>inducer</th>
<th>Plasmid</th>
<th>SEQ ID NO.:</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMPX-5</td>
<td>rhaRS</td>
<td>Rhamnos</td>
<td>pUC-18</td>
<td>6</td>
</tr>
<tr>
<td>pMPX-7</td>
<td>uidR</td>
<td>β-glucuronate</td>
<td>pUC-18</td>
<td>10</td>
</tr>
<tr>
<td>pMPX-8</td>
<td>melR</td>
<td>Melibiose</td>
<td>pUC-18</td>
<td>11</td>
</tr>
<tr>
<td>pMPX-18</td>
<td>araC</td>
<td>Arabinoose</td>
<td>pUC-18</td>
<td>12</td>
</tr>
<tr>
<td>pMPX-6</td>
<td>araC</td>
<td>Arabinoose</td>
<td>pUC-18</td>
<td>13</td>
</tr>
</tbody>
</table>

TABLE 14

<table>
<thead>
<tr>
<th>Oligonucleotide Primer Sequences for Table 13 Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRQ</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>69</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>71</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>73</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>77</td>
</tr>
<tr>
<td>78</td>
</tr>
<tr>
<td>79</td>
</tr>
<tr>
<td>80</td>
</tr>
<tr>
<td>81</td>
</tr>
<tr>
<td>82</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>84</td>
</tr>
<tr>
<td>85</td>
</tr>
<tr>
<td>86</td>
</tr>
</tbody>
</table>

[1057] Oligonucleotides SEQ ID NO.: 69, 70, 71 and 72 were used to amplify the rhaRS genes and their divergent
control region from the E. coli chromosome. Once amplified, this region was inserted into pUC18 using HindIII and PstI to create SEQ ID NO.: 6.

[1059] Oligonucleotides SEQ ID NOS.:73, 74, 75 and 76 were used to amplify the uidR control region, the uidR gene and the control region for expression from the E. coli chromosome. Once amplified, this region was inserted into pUC18 using HindIII and PstI to create SEQ ID NO.: 10.

[1060] Oligonucleotides SEQ ID NOS.:77, 78, 79 and 80 were used to amplify the mcrR gene and its divergent control region from the E. coli chromosome. Once amplified, this region was inserted into pUC18 using HindIII and SalI to create SEQ ID NO.: 11.

[1061] Oligonucleotides SEQ ID NOS.:81, 82, 83 and 84 were used to amplify the araC gene and its divergent control region from the E. coli chromosome. Once amplified, this region was inserted into pUC18 using HindIII and PstI to create SEQ ID NO.: 12.

[1062] Oligonucleotides SEQ ID NOS.:81, 82, 85 and 86 were used to amplify the araC gene and its divergent control region was PCR amplified with pBAD-24. Once amplified, this region was inserted into pEGFP (Clontech) using XhoI and SstI to create SEQ ID NO.: 13.

[1063] Except of pMPX-6, these expression constructs contain the same multiple cloning site. Therefore, any protein of interest may be inserted in each modular expression construct for simple expression screening and optimization.

[1064] By way of non-limiting example, other proteins that may be expressed are listed in Table 15.

### Table 15

<table>
<thead>
<tr>
<th>Protein</th>
<th>Origin</th>
<th>Construct</th>
<th>Purpose</th>
<th>SEQ ID NO.:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edg3</td>
<td>Rat</td>
<td>native</td>
<td>GPCR</td>
<td>14</td>
</tr>
<tr>
<td>β2AR</td>
<td>Human</td>
<td>native</td>
<td>GPCR</td>
<td>15</td>
</tr>
<tr>
<td>TNFR-1a</td>
<td>Human</td>
<td>residues</td>
<td>Receptor</td>
<td>18</td>
</tr>
<tr>
<td>(human)</td>
<td></td>
<td>29-455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFR-1b</td>
<td>Human</td>
<td>residues</td>
<td>Receptor</td>
<td>17</td>
</tr>
<tr>
<td>(human)</td>
<td></td>
<td>41-455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>Human</td>
<td>native</td>
<td>Gene</td>
<td>19</td>
</tr>
<tr>
<td>(human)</td>
<td></td>
<td></td>
<td>transfer</td>
<td></td>
</tr>
<tr>
<td>T-EGF</td>
<td>Human</td>
<td>chimeras</td>
<td>Gene</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>transfer</td>
<td></td>
</tr>
<tr>
<td>T-leuA</td>
<td>L.pseudotuberculosis</td>
<td>chimeras</td>
<td>Gene</td>
<td>21</td>
</tr>
</tbody>
</table>

### Table 16

<table>
<thead>
<tr>
<th>SEQ ID</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>GCAACACCAGCAAGAGCAGGGCCACC</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>CAATGTGATCTGATGATAGTCAGAGCC</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>GCGCGCATCGGCACACGCAGCAGGGCCACC</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>CCGCCGATCTCATTATCAGATGAGTCTGCA</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>GGGCAACCGCGGAGGACAGCGGCCC</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>CGACTGAGTCATTTGACTACAAATTTCC</td>
<td></td>
</tr>
</tbody>
</table>

[1065] Oligonucleotides SEQ ID NOS.:87, 88, 89 and 90 were used to amplify rat Edg3 from rat cDNA. Once amplified, this region was inserted into SEQ ID NO.: 6 (pMPX-5) using SalI and KpnI to create SEQ ID NO.: 14.
Oligonucleotides SEQ ID NOS: 91, 92, 93 and 94 were used to amplify human β2 adrenergic receptor (β2AR) from human heart cDNA. Once amplified, this region was inserted into SEQ ID NO: 6 (pMPX-5) using SalI and BamHI to create SEQ ID NO: 15.

Oligonucleotides SEQ ID NOS: 95, 96, 97 and 98 were used to amplify human tumor necrosis factor receptor (TNFR) residues 29-455 from human Jurkat CL71 cDNA. Once amplified, this region was inserted into SEQ ID NO: 12 (pMPX-18) using Sall and Kpnl to create SEQ ID NO: 18.

Oligonucleotides SEQ ID NOS: 99, 100, 101 and 102 were used to amplify human tumor necrosis factor receptor (TNFR) residues 41-455 from human Jurkat CL71 cDNA. Once amplified, this region was inserted into pBAD24 using Neol and XbaI to create SEQ ID NO: 17.

Oligonucleotides SEQ ID NOS: 103, 104, 105 and 106 were used to amplify human tumor necrosis factor (TNF) from human Jurkat CL71 cDNA. Once amplified, this region was inserted into SEQ ID NO: 13 (pMPX-6) using EcoRI and HindIII to create SEQ ID NO: 19.

| TABLE 17 |
| PROGRAM TO ANNEAL GRADIENT PCR WITH PfX POLYMERASE |

<table>
<thead>
<tr>
<th>Step</th>
<th>Temp (°C)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>95</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>Goto 2, 2X</td>
<td>0.5</td>
</tr>
<tr>
<td>6</td>
<td>68</td>
<td>0.5</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>68</td>
<td>2.5</td>
</tr>
<tr>
<td>9</td>
<td>Goto 6, 4X</td>
<td>0.5</td>
</tr>
<tr>
<td>10</td>
<td>95</td>
<td>0.5</td>
</tr>
<tr>
<td>11</td>
<td>60</td>
<td>0.5</td>
</tr>
<tr>
<td>12</td>
<td>68</td>
<td>2.5</td>
</tr>
<tr>
<td>13</td>
<td>Goto 10, 6X</td>
<td>0.5</td>
</tr>
<tr>
<td>14</td>
<td>95</td>
<td>0.5</td>
</tr>
<tr>
<td>15</td>
<td>58</td>
<td>0.5</td>
</tr>
<tr>
<td>16</td>
<td>68</td>
<td>2.5</td>
</tr>
<tr>
<td>17</td>
<td>Goto 14, 24X</td>
<td>hold</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>hold</td>
</tr>
<tr>
<td>19</td>
<td>end</td>
<td>hold</td>
</tr>
</tbody>
</table>

Oligonucleotides SEQ ID NOS: 107, 108, 109 and 110 were mixed and PCR amplified using anneal gradient PCR (Table 17) to form mature human epidermal growth factor (EGF) (residues 971-1023) translationally fused to the transmembrane domain of toxR from Vibrion cholerae. Once amplified, this region was inserted into SEQ ID NO: 13 (pMPX-6) using Kpnl and HindIII to create SEQ ID NO: 20.

Using PfX polymerase (Invitrogen) oligonucleotide SEQ ID NO: 111, 112, 113 and 114 were used to amplify invasin residues 490-986 (inv) from Yersinia pseudotuberculosis chromosomal DNA and form a translational fusion between the transmembrane domain of toxR from Vibrio cholerae. Once amplified, this region was inserted into SEQ ID NO: 13 (pMPX-6) using EcoRI and PstI to create SEQ ID NO: 21.

These proteins were proof-of-principle constructs used to evaluate the minicell platform. For purposes of this initial evaluation, all proteins except TNF, T-EGF and T-Invasin were cloned into pMPX-5, with these later proteins cloned into pMPX-6 for gene transfer experiments.

Whether the approach for protein expression is co-expression with minicell induction or expression following minicell and/or protoplast isolation, the procedure to transform the expression constructs is the same. To accomplish this, protein constructs were initially cloned into E. coli MG1655 and then into the minicell producing strain of interest. Transformation events were selected prior to minicell induction. For co-induction of protein and minicells, see the protocol for phoA expression above. For post-minicell and/or protoplast purification induction experiments, following minicell purification and/or protoplast preparation and purification, these cellular bodies were induced for protein production in either LBD or MDT at a minicell or protoplast/volume ratio of 1×10⁵ minicells or protoplasts/1 ml media. Media was supplemented with the appropriate inducer concentration (see Table 6). Protein induction is sensitive to a variety of factors including, but not limited to aeration and temperature, thus reaction volume to surface area ratio is important, as is the method of shaking and temperature of induction. Therefore, each protein must be treated as required to optimize expression. In addition to expression parameters, protoplasted minicells are sensitive to osmotic and mechanical forces. Therefore, protoplast protein induction reactions must also contain 10% sucrose with greater volume to surface area ratios than required for intact minicells to achieve similar aeration at lower revolutions.

Using the T-PhoA as a non-limiting example, protein expression was performed during and following minicell isolation. To accomplish this task, t-PhoA co-expressed with minicell induction was compared to t-PhoA expressed after minicell isolation. In both cases, overnight minicell-producing parental strains containing pMPX-5::t-PhoA were sub-cultured into LBD supplemented with the appropriate antibiotic. Cultures were grown to OD₅₀₀ 0.1 and induced for minicell production alone or for both minicell and protein production. Both cultures were harvested at OD₅₀₀ 1.0 and minicells produced were harvested as described above. Minicells to be induced for T-PhoA production following purification were induced by introducing 1×10⁷ purified minicells into a 15 ml culture tube containing 1 ml MDT with 1 mM L-rhamnose. Minicell protein induction was allowed to proceed for up to 14 hours and compared to protein production obtained using the co-expression approach. For each approach, minicells were fractionated and analyzed for membrane association, total protein, and membrane association-dependent enzymatic activity. These observations were compared to post-induction, pre-isolation parental cell/minicell (PC/MC) mixtures from the co-expressed reactions. The first observation was that co-expression of minicell and protein induction was superior to post-minicell purification induction (Table 18). Although the kinetics are slower for the post-minicell purification induction protocol, the end result is equivalent.

| TABLE 18 |
| COMPARATIVE PRODUCTION: CO-EXPRESSION VERSUS POST MINICELL PURIFICATION INDUCTION |

<table>
<thead>
<tr>
<th>Time of induction</th>
<th>Purified minicell induction</th>
<th>Co-expression induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>8.0</td>
<td>8.122</td>
</tr>
<tr>
<td>2.0</td>
<td>—</td>
<td>8.122</td>
</tr>
</tbody>
</table>
TABLE 18-continued

<table>
<thead>
<tr>
<th>Time of induction</th>
<th>Purified minicell induction</th>
<th>Co-expression induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>70.0</td>
<td>—</td>
</tr>
<tr>
<td>14.0</td>
<td>445.0</td>
<td>—</td>
</tr>
</tbody>
</table>

* Nanogram expressed T-PhoA per $1 \times 10^9$ minicells.

[1075] Using the co-expression induction procedure, the amount of membrane-associated T-PhoA was measured and compared for both parental cells and minicells. Briefly, following co-expression induction of T-PhoA and minicells, minicells were purified and their membranes isolated. For membrane isolation, minicells containing expressed T-PhoA were subjected to three rounds of freeze-thaw lysis in the presence of 10 μg/ml lysozyme. Following freeze-thaw cycling, the reaction was subjected to sonication. Sonicated material was centrifuged at 6,000 rpm in a microcentrifuge for 5 min at room temperature. Supernatants were transferred to a fresh 1.5 ml Eppendorf tube and centrifuged at 70,000 rpm using a TL-A100 rotor. Following centrifugation, the pellet was resuspended in buffer and analyzed for total T-PhoA protein (Table 19) and T-PhoA enzyme activity (Table 20).

TABLE 19

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Protein total</th>
<th>T-PhoA total</th>
<th>T-PhoA % total</th>
<th>Protein membrane associated</th>
<th>T-PhoA membrane associated</th>
<th>T-PhoA % membrane protein total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental cells</td>
<td>107.5</td>
<td>5.3</td>
<td>4.9</td>
<td>1.7</td>
<td>3.1</td>
<td>29.0</td>
</tr>
<tr>
<td>Minicells</td>
<td>4.6</td>
<td>0.8</td>
<td>17.5</td>
<td>1.0</td>
<td>0.5</td>
<td>50.0</td>
</tr>
<tr>
<td>Minicell IQ</td>
<td>25.2</td>
<td>4.4</td>
<td>5.5</td>
<td>2.7</td>
<td>5.5</td>
<td>—</td>
</tr>
</tbody>
</table>

* Total protein as determined by BCA assay (Pierce).
* Nanogram expressed T-PhoA per $1 \times 10^9$ minicells as determined via Western using an anti-PhoA antibody (Sigma) versus a PhoA standard curve (BCA determined).
* Equivalent membrane lipid to parental cell.

TABLE 20

| Protein ENZYMATIC ACTIVITY * (RELATIVE UNITS): PARENTAL CELLS VERSUS MINICELLS. |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Cell type | Unlysed | Lysed, total | Lysed, membrane |
| Parental cell | — | 358 | 240 |
| Minicell | 275 | 265 | 211 |
| Minicell IQ | 1,504 | 1,447 | 1,154 |

* Activity determined colorimetrically using PNP measuring optical density at 405 nm.
* Based on $1 \times 10^9$ parental cells or minicells per reaction.
* Equivalent membrane lipid to parental cell.

[1076] These results suggest that co-expression induction of T-PhoA and minicells together results in minicells containing an equivalent amount of T-PhoA produced in both parental cells and minicells. However, the percent of T-PhoA compared to total protein is 3.5x greater in minicells than in parental cells. Furthermore, of the protein made, T-PhoA constitutes 50% of the total membrane protein in minicells, whereas it is only 29% in parental cells. It should be noted that the T-PhoA protein associated with the membrane can be easily removed by treatment with mild, non-ionic detergent suggesting that the T-PhoA present in the membrane pellet is indeed associated with the membrane and not an insoluble, co-sedimenting precipitate (data not shown). Finally, PhoA is a periplasmic enzyme that requires export to the periplasmic space for proper folding and disulfide bond formation. Both of which are required for enzymatic activity. In the time course of this experiment, expression of ΔPhoA lacking a leader sequence does not demonstrate enzymatic activity. Furthermore, there is no difference between unlysed and lysed minicells containing expressed T-PhoA (Table 20) also demonstrating that the PhoA enzyme domain of the T-PhoA chimera must be present in the periplasmic space. Therefore, the T-PhoA construct must membrane associate and the PhoA domain must orient into the periplasmic space for enzymatic activity. Thus, when comparing equivalent amounts of membrane lipid between parental cells and minicells in Table 20, membrane association-dependent T-PhoA activity is almost 5x greater than in parental cells. Taking into account the data in Table 19 where 50% of T-PhoA is in the membrane compared to 29% in parental cells, the difference in T-PhoA membrane association is not sufficient to explain the almost 5x increase in minicell activity. These observations suggest that minicells contain a capacity to support more expressed membrane protein than parental cells and that the protein that associates with the membrane is more active. This activity may be simply result from minicells allowing greater efficiency of folding and disulfide bond formation for this particular protein. However, do to the fact that minicells do not contain chromosome, it is also possible that the overexpression of this protein is readily finding membrane-binding sites in the absence of chromosomally produced competitors present in parental cells. Furthermore, overexpression of proteins often leads to increased protease expression. Because minicells do not contain chromosome, these otherwise degraded surplus T-PhoA is allowed the continued opportunity to insert and properly fold in the membrane, an attribute that could lend favor to overexpression of more complex membrane proteins.

Example 18

Exemplary Methods to Induce and Study Complex Membrane Proteins

[1077] Expression of non-native (exogenous) complex membrane proteins in bacterial systems can be difficult. Using the minicell system, we are able to eliminate toxicity...
issues. However, issues still remain with proper translation, compartmentalization at the membrane, insertion in the membrane and proper folding for native activity. To account for these potential problems we have constructed a modular chimeric system that incorporates leader sequences and chaperone-recognized soluble domains that are native to our bacterial minicell system. In addition, we created constructs that overexpress the native chaperones groEL, and trigger factor (tig). Finally, we have constructed minicell-producing strains that contain mutations that effect protein export and disulfide bond formation. For non-limiting examples of these constructs see Table 21.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Residues of sequence</th>
<th>Purpose</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMPX-5:phoA leader</td>
<td>1-48</td>
<td>Membrane targeting</td>
<td>22</td>
</tr>
<tr>
<td>pMPX-5:phoA leader</td>
<td>1-49</td>
<td>Membrane targeting</td>
<td>23</td>
</tr>
<tr>
<td>pMPX-5:malE leader</td>
<td>1-28</td>
<td>Membrane targeting</td>
<td>24</td>
</tr>
<tr>
<td>pMPX-5:malE leader</td>
<td>1-370</td>
<td>Membrane targeting</td>
<td>25</td>
</tr>
<tr>
<td>pMPX-17 (groEL, tig)</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>pMPX-3:maElA:FLAG</td>
<td>2</td>
<td>Solubility</td>
<td>27</td>
</tr>
</tbody>
</table>

*Residues do not include FLAG sequence.

References to Table 21.

**TABLE 22**
OLIGONUCLEOTIDE PRIMER SEQUENCES FOR TABLE 21 CONSTRUCTS

<table>
<thead>
<tr>
<th>SEQ ID NO.</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>115</td>
<td>PhoA lead-1</td>
<td>GTCAGCGGCGAAGCTTATATAGCTGCC</td>
</tr>
<tr>
<td>116</td>
<td>PhoA lead-2</td>
<td>GGTGTCCGGGCTTGTGCTACAGG</td>
</tr>
<tr>
<td>117</td>
<td>PhoA lead-2 CCGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>PhoA lead-2-TG</td>
<td>CCAGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>119</td>
<td>PhoA complete</td>
<td>CAGCCGCCGAGCGGCGTCAGAGCC</td>
</tr>
<tr>
<td>120</td>
<td>PhoA complete-2XbaI</td>
<td>CAGCCGCCGAGCGGCGTCAGAGCC</td>
</tr>
<tr>
<td>121</td>
<td>MalE lead-1</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>122</td>
<td>MalE lead-2</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
</tbody>
</table>

**TABLE 22-continued**
OLIGONUCLEOTIDE PRIMER SEQUENCES FOR TABLE 21 CONSTRUCTS

<table>
<thead>
<tr>
<th>SEQ ID NO.</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>MalE-1</td>
<td>GGTGCACACCATCCCTCCTATCCGCC</td>
</tr>
<tr>
<td>124</td>
<td>MalE-2</td>
<td>CCGCATACCGCAAGAAGCCGATACCC</td>
</tr>
<tr>
<td>125</td>
<td>MalE-1-PstI</td>
<td>CCCGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>126</td>
<td>MalE-2-XbaI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>127</td>
<td>Tig-1</td>
<td>CCGGCACACCATCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>128</td>
<td>Tig-2</td>
<td>CCGGCACACCATCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>129</td>
<td>MalE-1-NarI</td>
<td>CCCGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>130</td>
<td>MalE-2-XbaI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>131</td>
<td>Gro-1</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>132</td>
<td>Gro-2</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>133</td>
<td>Gro-1-XbaI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>134</td>
<td>Gro-2-XbaI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>135</td>
<td>TrxA-1</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>136</td>
<td>TrxA-2</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>137</td>
<td>TrxA-1-PstI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
<tr>
<td>138</td>
<td>TrxA-2-FLAG-2XbaI</td>
<td>CGGCGGTCGCGGATCATCGGCAGGACGATTTAAGCTGCC</td>
</tr>
</tbody>
</table>

[1078] Oligonucleotides SEQ ID NO.:115, 116, 117 and 118 were used to amplify the phoA leader (residues 1-49) from E. coli chromosomal DNA. Once amplified, this region was inserted into SEQ ID NO.: 6 (pMPX-5) using PstI and XbaI to create SEQ ID NO.: 22.

[1079] Oligonucleotides SEQ ID NO.:115, 117, 119 and 120 were used to amplify the complete phoA gene from E. coli chromosomal DNA. Once amplified, this region was inserted into SEQ ID NO.: 6 (pMPX-5) using PstI and XbaI to create SEQ ID NO. 23.

[1080] Oligonucleotides SEQ ID NO.:121 and 122 were used to construct the malE leader (residues 1-28) sequence. Once annealed, this construct was inserted into SEQ ID NO.: 6 (pMPX-5) using PstI and XbaI to create SEQ ID NO.: 24.

[1081] Oligonucleotides SEQ ID NO.:123, 124, 125 and 126 were used to amplify the malE expanded leader (residues 1-370) from E. coli chromosomal DNA. Once amplified, this region was inserted into SEQ ID NO.: 6 (pMPX-5) using PstI and XbaI to create SEQ ID NO.: 25.

[1082] Oligonucleotides SEQ ID NO.:127, 128, 129 and 130 were used to amplify the tig control and gene region from E. coli chromosomal DNA. Once amplified, this region was
ligated to the groESL amplified region below using Xbali prior to insertion into SEQ ID NO.: 6 (pMPX-5) using Narl (from the tig region) and HindIII (from the groESL region) to create SEQ ID NO.: 26.

[1083] Oligonucleotides SEQ ID NOS.: 131, 132, 133 and 134 were used to amplify the groESL control and gene region from E. coli chromosomal DNA. Once amplified, this region was ligated to the amplified region above using Xbali prior to insertion into SEQ ID NO.: 6 (pMPX-5) using Narl (from the tig region) and HindIII (from the groESL region) to create SEQ ID NO.: 26.

[1084] Oligonucleotides SEQ ID NOS.: 135, 136, 137 and 138 were used to amplify trxA (residues 2-109) from E. coli chromosomal DNA and insert FLAG and Factor Xa sequences. Once amplified, this region was inserted into SEQ ID NO.: 6 (pMPX-5) using PstI and BamHI to create SEQ ID NO.: 27.

[1085] By way of non-limiting example, the pMPX-5: phoA leader (residues 1-48), pMPX-5: phoA leader (residues 1-494), pMPX-5: malE leader (residues 1-28), and pMPX-5: malE leader (residues 1-370) constructs are designed to direct expressed exogenous membrane proteins to the mini-cell cytoplasmic membrane. In addition to these constructs, By way of non-limiting example, mutations in E. coli genes secA and seeY, specifically mutation pPrA4 (Strader, J., et al. 1986. Kinetic analysis of lamB mutants suggest the signal sequence plays multiple roles in protein export. J. Biol. Chem. 261:15075-15080), permit promiscuous targeting to the membrane. These mutations, like the above constructs are integrated into the mini-cell expression system. To complement these mutations, the chaperone complex groESL and trigger factor have also been incorporated into the expression system. By way of non-limiting example, pMPX-5: trxA::FLAG will be used to create a carboxy-terminal fusion to the protein of interest to increase the membrane insertion efficiency of the membrane protein of interest (Tucker, J., and R. Grisshammer. 1996. Purification of a rat neurotransin receptor expressed in Escherichia coli. Biochem. J. 317:891-899). Also By way of non-limiting example, pMPX-5:FLAG::toxR and pMPX-5:FLAG::λcI constructs will be prepared to create a carboxy-terminal fusion to the protein of interest for use in a reporter-based assay for protein-protein interactions. By way of non-limiting example, the protein of interest for this system is a GPCR. Also By way of non-limiting example, this GPCR may be the neurotransin receptor from rat (Grisshammer, K., et al. 1993. Expression of a rat neurotransin receptor in Escherichia coli. Biochem. J. 295:571-576), or the β2 adrenergic receptor from humans (Freiismuth, M., et al. 1991. Expression of two β-adrenergic receptors in Escherichia coli: functional interaction with two forms of the stimulatory G protein. Proc. Natl. Acad. Sci. 88:8548-8552). Insertion of a GPCR into one of these reporter constructs creates a carboxy-terminal fusion between the GPCR of interest and the DNA-binding regulatory domain of the ToxR/positive activator, the λcI repressor, or the AraC positive activator. To complete this reporter system, By way of non-limiting example pMPX-5: (X)::toxR or pMPX-5: (X)::λcI will be used to create a carboxy-terminal fusion to the protein of interest for use in a reporter-based assay for protein-protein interactions, where (X) may be any protein or molecule involved in an intracellular or intermolecular interaction. By way of non-limiting example, this molecule of interest may be a G protein. This G protein may be the Gαs or Gαt protein from rat (Grisshammer, R., and E. Hermans. 2001. Functional coupling with Gq and G11 protein subunits promote high-affinity agonist binding to the neurotransin receptor NTS-1 expressed in Escherichia coli. FEBS Lett. 493:101-105), or the Gαs protein from human (Freiismuth, M., et al. 1991. Expression of two β-adrenergic receptors in Escherichia coli: functional interaction with two forms of the stimulatory G protein. Proc. Natl. Acad. Sci. 88:8548-8552). Like the GPCR, insertion of a G protein into one of these reporter constructs creates a carboxy-terminal fusion between the G protein of interest and the DNA-binding regulatory domain of the ToxR/positive activator, the λcI repressor, or other regulatory protein. Finally, these plasmid constructs contain the DNA-binding domain of each regulator; the ctx regulatory region from Vibrio cholerae (Russ, W. P., and D. M. Engelman. 1990. TOXCAT: a measure of transmembrane helix association in a biological membrane. 96:863-868), or the P21Og1 region of bacteriophage lambda (Hu, J. C., et al. 1990. Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. Science. 250:1400-1403), respectively. By way of non-limiting example, each binding domain is coupled to a reporter sequence encoding luciferase (Dunlap, P. V., and E. P. Greenberg. 1988. Control of Vibrio fischeri lux gene transcription by a cyclic AMP receptor protein-luxR protein regulatory circuit. J. Bacteriol. 170:4040-4046), green fluorescent protein (GFP) (Yang, T. T., et al. 1996. Dual color microscopic imaging of cells expressing the green fluorescent protein and a red-shifted variant. Gene. 173:19-23, Matthysse, A. G., et al. 1996. Construction of GFP vectors for use in gram-negative bacteria other than Escherichia coli. FEMS Microbiol. Lett. 145:87-94), or other reporter. Co-expression of these GPCRs and G protein chimeras will create a system measuring the interaction between a GPCR and G protein within an intact minicell. This system is designed to be used as a positive or negative read-out assay and may be used to detect loss or gain of GPCR function. Although the GPCR::G protein interaction is provided as an example, this modular system may be employed with any soluble or membrane protein system measuring protein-protein or other intermolecular interaction.

Example 19

Exemplary Methods for Gene Transfer Using Mini-cells or Minicell Protoplasts

[1086] Included in the design of the invention is the use of minicells to transfer genetic information to a recipient cell. By way of non-limiting example, this gene transfer may occur between a minicell and a mammalian cell in vitro, or in vivo, and this gene transfer may occur through cell-specific interactions, through general interactions, or a combination of each. To accomplish this task three basic constructs were created. Each of these constructs is created in pMPX-6 which contains a CMV promoter controlling the synthesis of GFP. The plasmid pMPX-6 was constructed by cloning the araC through the multiple cloning site of pBAD24 into pEGFP (Clontech). This construct provided a bacterial regulator as well as a method to monitor the success of gene transfer using GFP expression form the CMV promoter. In design, the protein expressed using the bacterial promoter will drive the
cell-cell interaction, while the successful transfer of DNA from the minicell to the recipient cell will initiate the production of GFP. By way of non-limiting example, proteins that will drive the cell-cell interaction may be the invasin protein from *Yersinia pseudotuberculosis*, which stimulates β1 integrin-dependent endocytic events. To properly display the invasin protein on the surface of minicells, the domain of invasin that stimulates these events (residues 490-986) (Dersch, P., and R. R. Isberg. 1999. A region of the *Yersinia pseudotuberculosis* invasin protein enhances integrin-mediated uptake into mammalian cells and promotes self-association. EMBO J. 18:1199-1213) was fused to the transmembrane domain of ToxR. Expression of this construct from pMPX-6 will display T-Inv on the surface of the minicell and stimulate endocytosis with any cell displaying a β1 integrin. Thus, T-Inv display will provide a general mechanism of gene transfer from minicells. To provide specificity, by way of non-limiting example, the ligand portion of epidermal growth factor (EGF) may be fused to the transmembrane domain of ToxR, thus creating a protein that will interact with cells displaying the EGF receptor (EGFR). Likewise, tumor necrosis factor (TNF) may also serve this purpose by stimulating cell-cell interactions between minicells displaying TNF and cells displaying TNF receptor (TNFR). Although EGF-EGFR and TNF-TNFR interactions may stimulate cell-cell fusion between minicells and recipient cells, or minicell uptake, this alone may not be sufficient to efficiently transfer genetic information from minicells. Therefore, a genetic approach to increasing the cell-cell genetic transfer may be the development of a genetic switch that senses the specificity interaction, e.g. EGF-EGFR interaction, and turns on the production of a second gene product, e.g. invasin, that stimulates the endocytic event. By way of non-limiting example, this genetic switch may be similar to the GPCR-G-protein interaction reporter system above, in that an extracellular event stimulates the dimerization of a transcriptional activator, thus turning on the production of invasin or invasin-like protein. In either approach, the display system to stimulate transfer of genetic information from minicells to recipient cells may also be applicable to the transfer of substances other than genetic information, e.g. pre-synthesized therapeutic drugs.

Expression in minicells may occur following purification of minicells and/or minicell protoplasts from parental cells and LPS constituents, respectively. However, for some applications it is preferred to co-express proteins of interest with minicell induction. For these approaches, one may use the protocol described in Example 13 for expression of the phoA constructs. Either of these approaches may be accomplished using one or more of the following expression constructs (Table 23) and/or optimized expression constructs (Table 25).

Expression plasmid pCGV1 contains a temperature sensitive lambda cl repressor (cl857) and both lambda PR and PL promoters (Guzman, C. A., et al. 1994. A novel *Escherichia coli* expression-export vector containing alkaline phosphatase as an insertional inactivation screening system. Gene. 148:171-172) with an atpE initiation region (Schauer, B., et al. 1987. Inducible expression vectors incorporating the *Escherichia coli* atpE translational initiation region. Gene. 52:279-283). Included in the design of the invention is the modification of this expression vector to best align the required Shine-Delgarno ribosomal binding site with cloning sites. In addition, the pCGV1 expression vector was modified to incorporate a stem-loop structure at the 3-prime end of the transcript in order to provide a strong transcriptional stop sequence (Table 23).

Expression plasmid pCI478 contains a temperature sensitive lambda cl repressor (cl857) and both lambda PR and PL promoters (Love, C. A., et al. 1996. Stable high-copy bacteriophage promoter vectors for overproduction of proteins in *Escherichia coli*. Gene. 176:49-53). Included in the design of the invention is the modification of this expression vector to best align the required Shine-Delgarno ribosomal binding site with cloning sites. In addition, the pCI478 expression vector was modified to incorporate a stem-loop structure at the 3-prime end of the transcript in order to provide a strong transcriptional stop sequence (Table 23).
TABLE 23

<table>
<thead>
<tr>
<th>New plasmid</th>
<th>Parent plasmid</th>
<th>Region removed</th>
<th>Region added</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMPX-84</td>
<td>pCGV1</td>
<td>NdeI - BamHI</td>
<td>NdeI, SD - PstI</td>
<td>139</td>
</tr>
<tr>
<td>pMPX-85</td>
<td>pCGV1</td>
<td>NdeI - BamHI</td>
<td>NdeI, SD - SalI</td>
<td>140</td>
</tr>
<tr>
<td>pMPX-86</td>
<td>pCL478</td>
<td>BamHI - XhoI</td>
<td>BamHI, SD - PstI</td>
<td>141</td>
</tr>
<tr>
<td>pMPX-87</td>
<td>pCL478</td>
<td>BamHI - XhoI</td>
<td>SalI, XhoI, KpnI</td>
<td>142</td>
</tr>
</tbody>
</table>

*"SD" refers to a Shine-Delgarno ribosome-binding sequence; "Stem-loop" refers to a stem-loop structure that functions as a transcriptional stop site.

TABLE 24

<table>
<thead>
<tr>
<th>SEQ NO</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>143</td>
<td>CSV1-1 - Sall</td>
<td>TATGTAGAGAGTTCTGGACGCTCTAGAATGGAGCT</td>
</tr>
<tr>
<td>144</td>
<td>CSV1-2 - Sall</td>
<td>GATCTCAGACGCTCTAGAATGGAGCT</td>
</tr>
<tr>
<td>145</td>
<td>CSV1-1 - PstI</td>
<td>CCGCCGCTCTAGAATGGAGCT</td>
</tr>
<tr>
<td>146</td>
<td>CSV1-2 - PstI</td>
<td>AGCTAGAGGCCTCACAGACGCT</td>
</tr>
<tr>
<td>147</td>
<td>CSV1-1 - Sall</td>
<td>GATCTCAGACGCTCTAGAATGGAGCT</td>
</tr>
<tr>
<td>148</td>
<td>CSV1-2 - Sall</td>
<td>AGCTAGAGGCCTCACAGACGCT</td>
</tr>
<tr>
<td>149</td>
<td>CSV1-1 - PstI</td>
<td>CCGCCGCTCTAGAATGGAGCT</td>
</tr>
<tr>
<td>150</td>
<td>CSV1-2 - PstI</td>
<td>AGCTAGAGGCCTCACAGACGCT</td>
</tr>
</tbody>
</table>

[1091] Oligonucleotides SEQ ID NOS.: 143 and 144 were annealed to each other to generate a DNA molecule with a 5' overhang at both ends. The overhangs are designed so that the DNA can be directly cloned into pCGVI cut with NdeI (5' overhang is TA) and BamHI (5' overhang is GATC). Insertion of the annealed DNA into pCGVI creates SEQ ID NO.: 139, pMPX-84.

[1092] Oligonucleotides SEQ ID NOS.: 145 and 146 were annealed to each other to generate a DNA molecule with a 5' overhang at both ends. The overhangs are designed so that the DNA can be directly cloned into pCGVI cut with NdeI (5' overhang is TA) and BamHI (5' overhang is GATC). Insertion of the annealed DNA into pCGVI creates SEQ ID NO.: 140, pMPX-85.

[1093] Oligonucleotides SEQ ID NOS.: 147 and 148 were annealed to each other to generate a DNA molecule with a 5' overhang at both ends. The overhangs are designed so that the DNA can be directly cloned into pCL478 cut with BamHI (5' overhang is GATC) and XhoI (overhang is TCGA). Insertion of the annealed DNA into pCL478 cut with BamHI and XhoI creates SEQ ID NO.: 141, pMPX-86.

[1094] Oligonucleotides SEQ ID NOS.: 149 and 150 were annealed to each other to generate a DNA molecule with a 5' overhang at both ends. The overhangs are designed so that the DNA can be directly cloned into pCL478 cut with BamHI (5' overhang is GATC) and XhoI (overhang is TCGA). Insertion of the annealed DNA into pCL478 cut with BamHI and XhoI creates SEQ ID NO.: 142, pMPX-87.

The optimized expression constructs in Table 25 were created from SEQ ID NOS.: 6, 11, and 12 (see Table 13). Modifications were made to optimize the alignment of the SalI or PstI cloning sites with the Shine-Delgarno ribosome-binding site. In addition, stem-loop transcriptional termination sequences were added on the 3' end of the cloning region.

TABLE 25

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Regulatory element(s)</th>
<th>Plasmid</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMPX-67</td>
<td>RhaRS</td>
<td>PUC-18</td>
<td>151</td>
</tr>
<tr>
<td>pMPX-72</td>
<td>RhaRS</td>
<td>PUC-18</td>
<td>152</td>
</tr>
<tr>
<td>pMPX-66</td>
<td>AraC</td>
<td>PUC-18</td>
<td>153</td>
</tr>
<tr>
<td>pMPX-71</td>
<td>AraC</td>
<td>PUC-18</td>
<td>154</td>
</tr>
<tr>
<td>pMPX-68</td>
<td>MelR</td>
<td>PUC-18</td>
<td>155</td>
</tr>
</tbody>
</table>

TABLE 26

<table>
<thead>
<tr>
<th>SEQ NO</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>Rha-1</td>
<td>GCGGAATTGATAGGAGCCCGTG</td>
</tr>
<tr>
<td>156</td>
<td>Rha-SD</td>
<td>GCAGGAACCTCCAGAAATTTGACTAG</td>
</tr>
<tr>
<td>71</td>
<td>Rha-1</td>
<td>CGCGCAGCTCTAGACTGACGCT</td>
</tr>
<tr>
<td>157</td>
<td>Rha-SD</td>
<td>CGCGCAGCTCTAGACTGACGCT</td>
</tr>
<tr>
<td>158</td>
<td>Rha-SD</td>
<td>CCGCGGAGGGTACATGGGAGCGG</td>
</tr>
<tr>
<td>01</td>
<td>Ara-1</td>
<td>CCCGAGGATAGGAGCCCGTG</td>
</tr>
<tr>
<td>159</td>
<td>Ara-SD</td>
<td>CGAGGATAGGAGCCCGTG</td>
</tr>
<tr>
<td>83</td>
<td>Ara-1</td>
<td>GCAGGAACCTCAGAAATTTGACTAG</td>
</tr>
</tbody>
</table>

[1095] The optimized expression constructs in Table 25 were created from SEQ ID NOS.: 6, 11, and 12 (see Table 13). Modifications were made to optimize the alignment of the SalI or PstI cloning sites with the Shine-Delgarno ribosome-binding site. In addition, stem-loop transcriptional termination sequences were added on the 3' end of the cloning region.
TABLE 26-continued

| SEQ
<table>
<thead>
<tr>
<th>ID NO.</th>
<th>Primer name</th>
<th>5’ to 3’ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>Ara-SD</td>
<td>CCAGCGGATACCCCTTGACGCTTTCGAGTTGGGCG</td>
</tr>
<tr>
<td></td>
<td>SalI</td>
<td>CGGAATTCCTGCCGTAAG</td>
</tr>
<tr>
<td></td>
<td>KpnI</td>
<td>GCCTCGAGCAGTATG</td>
</tr>
<tr>
<td>161</td>
<td>Ara-SD</td>
<td>CCAGCGGATACCCCTTGACGCTTTCGAGTTGGGCG</td>
</tr>
<tr>
<td></td>
<td>PstI</td>
<td>CGGAATTCCTGCCGTAAG</td>
</tr>
<tr>
<td></td>
<td>KpnI</td>
<td>GCCTCGAGCAGTATG</td>
</tr>
<tr>
<td>77</td>
<td>Mel-1</td>
<td>CGCTTTTACCCGGAGAAG</td>
</tr>
<tr>
<td>162</td>
<td>Mel-1</td>
<td>CGCTTTTACCCGGAGAAG</td>
</tr>
<tr>
<td>79</td>
<td>Mel-1</td>
<td>GCCGCAAGCTCGTCATCAG</td>
</tr>
<tr>
<td>163</td>
<td>Mel-1</td>
<td>GCCGCAAGCTCGTCATCAG</td>
</tr>
<tr>
<td></td>
<td>HindIII</td>
<td>GCCGCAAGCTCGTCATCAG</td>
</tr>
<tr>
<td></td>
<td>KpnI</td>
<td>GCCTCGAGCAGTATG</td>
</tr>
</tbody>
</table>

[1096] Oligonucleotides SEQ ID NOS.: 69, 156, 72, and 157 were used to amplify the rhaRS genes and their divergent control region from the *E. coli* chromosome and insertion of an optimized Sall-Shine-Delgarno ribosome-binding alignment and a stem-loop transcriptional termination sequence. Once amplified, this region was inserted into pUC18 using HindIII and KpnI to create pMX-64. pMX-64, SEQ ID NO.: 154.

[1100] Oligonucleotides SEQ ID NOS.: 77, 162, 79, 163 were used to amplify the melR genes and their divergent control region from the *E. coli* chromosome and insertion of an optimized Sall-Shine-Delgarno ribosome-binding alignment and a stem-loop transcriptional termination sequence. Once amplified, this region was inserted into pUC18 using HindIII and KpnI to create, pMX-68, SEQ ID NO.: 155.

Example 21

Optimization of Rat Neurotensin Receptor (NTR) Expression


TABLE 27

<table>
<thead>
<tr>
<th>NEUROTENSIN RECEPTOR EXPRESSION FACILITATING CONSTRUCTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein a</td>
</tr>
<tr>
<td>MaIE(L) Sall-MaIE(1-370)-Factor Xa-NTR Ionom</td>
</tr>
<tr>
<td>NTR Factor Xa-NTR(43-424)-Nol-FLAG-KpnI</td>
</tr>
<tr>
<td>MaIE(L)-NTR Sall-MaIE(1-370)-Factor Xa-NTR(43-424)-Nol-FLAG-KpnI</td>
</tr>
<tr>
<td>MaIE(S)-NTR Sall-MaIE(1-28)-Factor Xa-NTR(43-424)-Nol-FLAG-KpnI</td>
</tr>
<tr>
<td>Txa NoT-1 NoT-2</td>
</tr>
<tr>
<td>MaIE(L)-NTR Txa Sall-MaIE(1-370)-Factor Xa-NTR(43-424)-Nol-FLAG-KpnI</td>
</tr>
<tr>
<td>MaIE(S)-NTR Txa Sall-MaIE(1-28)-Factor Xa-NTR(43-424)-Nol-FLAG-KpnI</td>
</tr>
</tbody>
</table>

a (L) refers to MaIE residues 1-370, and (S) refers to MaIE residues 1-28.
b All mature constructs were cloned into Sall and KpnI sites of SEQ ID NOS.: 140, 142, 151 and 153.

TABLE 28

<table>
<thead>
<tr>
<th>Oligonucleotide PRIMER SEQUENCES FOR TABLE 27</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ ID NO.</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>171</td>
</tr>
<tr>
<td>172</td>
</tr>
<tr>
<td>173</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>174</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>175</td>
</tr>
</tbody>
</table>

[1097] Oligonucleotides SEQ ID NOS.: 69, 156, 72, and 158 were used to amplify the rhaRS genes and their divergent control region from the *E. coli* chromosome and insertion of an optimized Pstl-Shine-Delgarno ribosome-binding alignment and a stem-loop transcriptional termination sequence. Once amplified, this region was inserted into pUC18 using HindIII and KpnI to create pMX-67, SEQ ID NO.: 151.

[1098] Oligonucleotides SEQ ID NOS.: 69, 156, 72, and 158 were used to amplify the rhaRS genes and their divergent control region from the *E. coli* chromosome and insertion of an optimized Pstl-Shine-Delgarno ribosome-binding alignment and a stem-loop transcriptional termination sequence. Once amplified, this region was inserted into pUC18 using HindIII and KpnI to create pMX-67, SEQ ID NO.: 152.

[1099] Oligonucleotides SEQ ID NOS.: 69, 156, 72, and 158 were used to amplify the rhaRS genes and their divergent control region from the *E. coli* chromosome and insertion of an optimized Pstl-Shine-Delgarno ribosome-binding alignment and a stem-loop transcriptional termination sequence. Once amplified, this region was inserted into pUC18 using HindIII and KpnI to create, pMX-64, SEQ ID NO.: 154.
Oligonucleotides SEQ ID NOS.: 172, 171, 172, 173 and 174 were used to amplify mafE residues 1-370 from the E. coli chromosome to create SEQ ID NO.: 164. Using overlap PCR with the extended NTR homology, a chimeric translational fusion was made between MafE (1-370) and NTR residues 43-424 (SEQ ID NO.: 165) to create a SEQ ID NO.: 166. SEQ ID NO.: 166 was cloned into plasmids pMPX-85, pMPX-87, pMPX-66 and pMPX-67 (respectively, SEQ ID NO.: 140, 142, 151 and 153) using Sall and Kpnl.

Three-step PCR with oligonucleotides, SEQ ID NOS.: 175 and 176 as primers was used to amplify NTR residues 43-424 from rat brain cDNA. SEQ ID NOS.: 177 and 178 were then used with the NTR (43-424) template to add factor Xa and FLAG sequence. Finally, SEQ ID NOS.: 177 and 179 were used to add a KpnI site to create SEQ ID NO.: 165. Using overlap PCR with mafE(1-370) containing extended NTR homology, a chimeric translational fusion was made between NTR (43-424) and MafE (1-370) (SEQ ID NO.: 164) to create a SEQ ID NO.: 166. SEQ ID NO.: 166 was cloned into SEQ ID NOS.: 140, 142, 151 and 153 using Sall and KpnI.

Using three-step PCR oligonucleotides SEQ ID NOS.: 175 and 176 were first used to amplify NTR residues 43-424 from rat brain cDNA. SEQ ID NOS.: 178 and 180 were then used with the NTR (43-424) template to add factor Xa and FLAG sequence. Finally, SEQ ID NOS.: 179 and 181 were used to add a KpnI site to create SEQ ID NO.: 167. SEQ ID NO.: 167 was cloned into SEQ ID NOS.: 140, 142, 151 and 153 using Sall and KpnI.

Methods for Functional GPCR Assay

Functional G-protein-coupled receptor (GPCR) binding assays in minicells requires expression of a GPCR of interest into the minicell membrane bilayer and cytoplasmic expression of the required G-protein. For these purposes, constructs were created to co-express both a GPCR and a G-protein. To regulate the ratio of GPCR to G-protein, translational fusions were created. In these constructs, the GPCR and G-protein are co-transcribed as a bi-cistronic mRNA. To measure the GPCR-G-protein interaction in the intact minicell, each protein was created as a chimera with a transactivation domain. For these studies the N-terminal DNA-binding, activation domain of the ToxR protein from V. cholerae was fused to the C-terminus of both the GPCR and G-protein. Finally, to measure the interaction GPCR-G-protein interaction, the ToxR-activated ctx promoter region was cloned in front of lacZ. Dimerization of the ToxR DNA-binding region will bind and activate the ctx promoter. In this construct, heterodimerization of the GPCR and G-protein will promote dimerization of ToxR that will be monitored by LacZ expression. Details of these constructs are shown in Table 29.
<table>
<thead>
<tr>
<th>SEQ ID NO.:</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>209</td>
<td>β2AR-1</td>
<td>GGGCACCACCGCGAACGCGGCC</td>
</tr>
<tr>
<td>210</td>
<td>β2AR-2</td>
<td>GCAGTGCGTCTGGTATCGTCCTCG</td>
</tr>
<tr>
<td>211</td>
<td>β2AR-1-1</td>
<td>GCGGCCGCG</td>
</tr>
<tr>
<td>212</td>
<td>β2AR-2-1</td>
<td>AGCCTTGCTAGGTATCAGTCCTTG</td>
</tr>
<tr>
<td>213</td>
<td>G51a-1</td>
<td>GGGCCTCGAGATTCAGACGAG</td>
</tr>
<tr>
<td>214</td>
<td>G51a-2</td>
<td>GACAGCGCTCTGATGGCATCAG</td>
</tr>
<tr>
<td>215</td>
<td>G51a-1-1</td>
<td>GAGGATGTTGACGGGACGAG</td>
</tr>
<tr>
<td>216</td>
<td>G51a-2-1</td>
<td>TGCGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>217</td>
<td>G51a-2-1-1</td>
<td>GGACGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>218</td>
<td>G51a-2-1-2</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>219</td>
<td>G51a-2-2</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>220</td>
<td>G51a-2-2-1</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>221</td>
<td>G51a-2-2-2</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>222</td>
<td>G51a-2-3</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>223</td>
<td>G51a-2-4</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>224</td>
<td>G51a-2-5</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
</tbody>
</table>

**TABLE 30-continued**

<table>
<thead>
<tr>
<th>SEQ ID NO.:</th>
<th>Primer name</th>
<th>5' to 3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>235</td>
<td>β2AR-2-2-1</td>
<td>CTCGAAATTTTATCTGACGTAG</td>
</tr>
<tr>
<td>236</td>
<td>Tox (S-141)-1</td>
<td>GAAGTACCCAGCAGAGACGAG</td>
</tr>
<tr>
<td>237</td>
<td>Tox (S-141)-2</td>
<td>GAAGTACCCAGCAGAGACGAG</td>
</tr>
<tr>
<td>238</td>
<td>Tox (S-141)-3</td>
<td>GAAGTACCCAGCAGAGACGAG</td>
</tr>
<tr>
<td>239</td>
<td>Tox (S-141)-4</td>
<td>GAAGTACCCAGCAGAGACGAG</td>
</tr>
<tr>
<td>240</td>
<td>Ctx-1</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>241</td>
<td>Ctx-2</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>242</td>
<td>Ctx-3</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>243</td>
<td>Ctx-4</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>244</td>
<td>Ctx-5</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>245</td>
<td>Ctx-6</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>246</td>
<td>Ctx-7</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
<tr>
<td>247</td>
<td>Ctx-8</td>
<td>CGCGGGCGCCGCGCCCGG</td>
</tr>
</tbody>
</table>

[1107] Oligonucleotides SEQ ID NOS.: 209, 210, 211 and 212 were used to amplify human β2AR from human cDNA to create SEQ ID NO.: 186. Using Sall and Xhol a translational fusion was made between β2AR and human GS1α (SEQ ID NO.: 187) to create a SEQ ID NO.: 188; SEQ ID NO.: 188 was cloned into SEQ ID NOS.: 140, 142, 151 and 153 using Sall and Xbal.

[1108] Oligonucleotides SEQ ID NOS.: 213, 214, 215 and 216 were used to amplify human GS1α from human cDNA to create SEQ ID NO.: 187. Using Xhol and Xbal a translational fusion was made between GS1α and human β2AR (SEQ ID NO.: 186) create SEQ ID NO.: 188. SEQ ID NO.: 188 was cloned into SEQ ID NOS.: 140, 142, 151 and 153 using Sall and Xbal.

[1109] Oligonucleotides SEQ ID NOS.: 213, 214, 215 and 217 were used to amplify human GS1α from human cDNA to create SEQ ID NO.: 192. Using Xhol and Xbal a translational fusion was made with ToxR residues 5-141 from *Vibrio cholerae* (SEQ ID NO.: 191) to create SEQ ID NO.: 197. To be used to create a transcriptional fusion with β2AR-ToxR chimeras as shown in SEQ ID NO.: 205 and future GPCR-ToxR chimeras.

[1110] Oligonucleotides SEQ ID NOS.: 218, 219, 220 and 221 were used to amplify human Go12/13 from human...
cDNA to create SEQ ID NO.: 196. Using XhoI and XbaI a translational fusion was made with ToxR residues 5-141 from Vibrio cholerae (SEQ ID NO.: 191) to create SEQ ID NO.: 201. To be used to create future transcriptional fusions with GPCR-ToxR chimeras as shown in SEQ ID NO.: 205.

[1111] Oligonucleotides SEQ ID NOS.: 222, 223, 224 and 225 were used to amplify human Gαq from human cDNA to create SEQ ID NO.: 194. Using XhoI and XbaI a translational fusion was made with ToxR residues 5-141 from Vibrio cholerae (SEQ ID NO.: 191) to create SEQ ID NO.: 199. To be used to create future transcriptional fusions with GPCR-ToxR chimeras as shown in SEQ ID NO.: 205.

[1112] Oligonucleotides SEQ ID NOS.: 226, 227, 228 and 229 were used to amplify human Gzα from human cDNA to create SEQ ID NO.: 195. Using XhoI and XbaI a translational fusion was made with ToxR residues 5-141 from Vibrio cholerae (SEQ ID NO.: 191) to create SEQ ID NO.: 200. To be used to create future transcriptional fusions with GPCR-ToxR chimeras as shown in SEQ ID NO.: 205.

[1113] Oligonucleotides SEQ ID NOS.: 230, 231, 232 and 233 were used to amplify human GS2α from human cDNA to create SEQ ID NO.: 193. Using XhoI and XbaI a translational fusion was made with ToxR residues 5-141 from Vibrio cholerae (SEQ ID NO.: 191) to create SEQ ID NO.: 198. To be used to create future transcriptional fusions with GPCR-ToxR chimeras as shown in SEQ ID NO.: 205.

[1114] Oligonucleotides SEQ ID NOS.: 209, 210, 211 and 234 were used to amplify human β2AR from human cDNA to create SEQ ID NO.: 189. Using Sall and XhoI a transcriptional fusion was made between β2AR and human GS1α (SEQ ID NO.: 187) to create a SEQ ID NO.: 190. SEQ ID NO.: 190 was cloned into SEQ ID NOS.: 140, 142, 151 and 153 using Sall and XbaI.

[1115] Oligonucleotides SEQ ID NOS.: 236, 237, 238 and 239 were used to amplify bases coinciding with ToxR residues 5-141 from Vibrio cholerae to create SEQ ID NO.: 202. Using PstI and XhoI a translational fusion was made between ToxR and human β2AR (SEQ ID NO.: 203) to create SEQ ID NO.: 204.

[1116] Oligonucleotides SEQ ID NOS.: 209, 210, 211 and 235 were used to amplify human β2AR from human cDNA to create SEQ ID NO.: 203. Using Sall and PstI a translational fusion was made between β2AR and ToxR (SEQ ID NO.: 202) to create SEQ ID NO.: 204.

[1117] Oligonucleotides SEQ ID NOS.: 197 and 204 transcriptional fusions were created between the β2AR-ToxR translational fusion (SEQ ID NO.: 204) and the GS1α-ToxR translational fusion (SEQ ID NO.: 197) to create SEQ ID NO.: 205.

[1118] Oligonucleotides SEQ ID NOS.: 240, 241, 242 and 243 were used to amplify the ctp promoter region (Pctp) from Vibrio cholerae to create SEQ ID NO.: 206. Combining this PCR product in combination with the SEQ ID NO.: 207 PCR product and amplifying in the presence of SEQ ID NOS.: 242, 247, SEQ ID NO.: 208 was created. Using Xbal, the SEQ ID NO.: 208 reporter construct was subsequently cloned into pACYC184 for co-transformation with the GPCR-G-protein fusions constructs above.

[1119] Oligonucleotides SEQ ID NOS.: 244, 245, 246 and 247 were used to amplify the lacZ from E. coli to create SEQ ID NO.: 207. Combining this PCR product in combination with the SEQ ID NO.: 206 PCR product and amplifying in the presence of SEQ ID NOS.: 242 and 247, SEQ ID NO.: 208 was created. Using Xbal, the 208 reporter construct was subsequently cloned into pACYC184 for co-transformation with the GPCR-G-protein fusions constructs above.

Example 23
Modular Membrane-Targeting and Solubilization Expression Constructs

[1120] To produce membrane proteins efficiently in mini-cells it may be necessary to create chimeric fusions with the membrane protein of interest. In this Example, various regions of the MalE protein have been cloned into a modular expression system designed to create chimeric fusions with direct difficult to target membrane proteins to produce leader domains that will direct the proteins to the cytoplasmic membrane (Miller, K., et al. 1998. Production of active chimeric pediocin AεII in Escherichia coli in the absence of processing and secretion genes from the Pediococcus pap operon. Appl. Environ. Microbiol. 64:14-20). Similarly, a modified version of the TrxA protein has also been cloned into this modular expression system to create chimeric fusions with proteins that are difficult to maintain in a soluble conformation (LaVallie, E.R., et al. 1993. A thymoexin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology (N.Y.) 11:187-193). Table 31 describes each of these modular constructs.

TABLE 31
MODULAR MEMBRANE-TARGETING AND SOLUBILIZATION EXPRESSION CONSTRUCTS

<table>
<thead>
<tr>
<th>Protein *</th>
<th>Construct *</th>
<th>SEQ ID NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>MalE (1-28)</td>
<td>N terminus-MalE(1-28)-Factor Xa-PstI, Sall, XbaI-FLAG, NheI</td>
<td>248</td>
</tr>
<tr>
<td>TrxA (2-109, del 103-107)</td>
<td>PstI, Sall, XbaI-TrxA(2-109, del 103-107)-FLAG-NheI</td>
<td>250</td>
</tr>
<tr>
<td>MalE (1-28)-TrxA (2-109, del 103-107)-FLAG-NheI</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td>MalE (1-370, del 354-364)-Factor Xa-PstI, Sall, XbaI-TrxA(2-109, del 103-107)-FLAG-NheI</td>
<td>252</td>
<td></td>
</tr>
</tbody>
</table>

* The term “del” refers to a deletion in which amino acid residues following the term “del” are removed from the sequence.
<table>
<thead>
<tr>
<th>SEQ ID Primer NO. name 5’ to 3’ sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>253 MalR-1- NhelI</td>
</tr>
<tr>
<td>254 MalR-2- NhelI</td>
</tr>
<tr>
<td>255 MalR-3- NhelI</td>
</tr>
<tr>
<td>256 MalR-4- NhelI</td>
</tr>
<tr>
<td>257 MalR-1a Gctgca</td>
</tr>
<tr>
<td>258 MalR-2a Gctgca</td>
</tr>
<tr>
<td>259 MalR-3a Gctgca</td>
</tr>
<tr>
<td>260 MalR-2b Gctgca</td>
</tr>
<tr>
<td>261 MalR-3b Gctgca</td>
</tr>
<tr>
<td>262 MalR-1a Ccggct</td>
</tr>
<tr>
<td>263 MalR-2a Ccggct</td>
</tr>
<tr>
<td>264 MalR-3a Ccggct</td>
</tr>
<tr>
<td>265 MalR-3b Ccggct</td>
</tr>
</tbody>
</table>

**Example 24**

Poroplast™ Formation

**[1121]** Oligonucleotides SEQ ID NOS.: 253, 254, 255 and 256 overlap with each other to form a scaffold template to PCR amplify maLE (1-28) to create a SEQ ID NO.: 248. Following PCR amplification, SEQ ID NO.: 248 was digested with Nhel and cloned into SEQ ID NOS.: 152, 154, 139 and 141 digested with PstI and XbaI. The resultant products create SEQ ID NOS.: 266, 267, 268 and 269, respectively, that lose both the 5’ PstI and 3’ PstI restriction sites and retain the PstI, Sall, and XbaI restriction sites between maLE (1-28) and the FL-agarase sequence. Insertion of a protein in alignment with these sites results in a chimeric protein containing amino-terminal maLE (1-370, del 354-364) and carboxy-terminal FL-agarase.

**Example 24**

Poroplast™ Formation

**[1127]** Minicells are used to prepare Poroplasts in order to increase the accessibility of a membrane protein component and/or domain to the outside environment. Membrane proteins in the inner membrane are accessible for lipid binding and/or other interactions in porplasts due to the absence of an outer membrane. The removal of the outer membrane from E. coli whole cells and minicells to produce poroplasts was carried out using modifications of previously described protoplast and analysis protocols (Birdsell et al., Production and Ultrastucture of Lysozyme and Ethylendiaminetetraetate-1-yosyme Spheroplasts of Escherichia coli, J. Bacteriol. 93:427-437, 1967; Weiss et al., Protoplast Formation in Escherichia coli, J. Bacteriol. 128:668-670, 1976; Matsuyama, S.-I., et al. SecO is involved in the release of translocated secretory proteins from the cytoplasmic membrane of Escherichia coli. 12:265-270, 1993).

**[1128]** In brief, cells were grown to late-log phase and pelleted at room temperature. Minicells were also isolated from cultures in late-log phase. The pellet was washed twice with 50 mM Tris, pH 8.0. Following the second wash, 1x10^6 cells were resuspended in 1 ml 50 mM Tris (pH 8.0) that contained 8% sucrose and 2 mM EDTA. Cell/EDTA/sucrose mixtures were incubated at 37°C for 10 min, centrifuged, decanted, and poroplasted cells were resuspended in 50 mM Tris, pH 8.0 with 8% sucrose. Incubation with anti-LPS-coated magnetic beads, as described in Example 14, is used to enrich for poroplasts that lack LPS. Following incubation
with the resuspended protoplasted cells, the anti-LPS magnetic beads were removed from suspension with a magnet.

To examine the range of molecular sizes that can pass through the cell wall, an IgG molecule was tested for its ability to pass the intact cell wall. Binding of an antibody to the ToxR-PhoA chimera expressed on the inner membrane minicell poroplasts was measured. Briefly, minicell poroplasts with and without inner membrane-bound ToxR-PhoA were incubated at 37° C, with anti-PhoA antibody in reaction buffer (50 mM Tris, pH 8.0, 8% sucrose, 1% BSA, and 0.01% Tween-20). Following incubation, poroplasts were centrifuged, washed 3 times with reaction buffer, and resuspended in 50 mM Tris, pH 8.0 with 8% sucrose. Following resuspension, bound proteins from 5x10^7 minicells or minicell poroplasts were separated using denaturing SDS-PAGE, transferred to nitrocellulose, and developed using both anti-PhoA antibody and secondary antibody against both heavy and light chains of anti-PhoA IgG (Table 33).

**TABLE 33**

<table>
<thead>
<tr>
<th>EDTA (mM)</th>
<th>0</th>
<th>2</th>
<th>0</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysozyme (ng/ml)</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protoplasts (ng antibody bound)</th>
<th>Poroplasts (ng antibody bound)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minicell-toxR-PhoA</td>
<td>ND *</td>
</tr>
<tr>
<td>Minicell only</td>
<td>ND *</td>
</tr>
</tbody>
</table>

* Non-detectable

These results demonstrate that the cell wall present on poroplasts is penetrable by an IgG molecule and that an IgG molecule is capable of passing the intact cell wall and binding to an inner membrane protein. From this data it appears that poroplast operate at ~10% the efficiency of protoplasts by allowing 0.6 ng of IgG to bind inner membrane-bound ToxR-PhoA compared to 12.8 ng. However, given the large size of IgG (~150,000 Daltons) it is expected that molecules having a smaller molecular weight will efficiently access inner membrane proteins in poroplasts.

Example 25

Production of Neurtensin Receptor (NTR)

To demonstrate expression of NTR in isolated minicells, MalE(L)-NTR (SEQ ID NO: 166) was cloned into pMP6-67 (SEQ ID NO: 151). Following minicell isolation, 1.5x10^8 minicells were induced with 1 mM Rhamnose for 2 hour at 37° C. Following induction, the protein produced was visualized via Western analysis using an anti-MalE antibody following separation on an SDS-PAGE. The results are shown in FIG. 2.

These data demonstrate that MalE(L)-NTR is induced 87-fold by addition of 1 mM rhamnose to the minicell induction mixture. Cross-reactive proteins are host MalE and non-specific binding by Goat-anti-mouse HRP secondary antibody.

The contents of the articles, patents, and patent applications, and all other documents and electronically available information mentioned or cited herein, are hereby incorporated by reference in their entirety to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference. Applicants reserve the right to physically incorporate into this application any and all materials and information from any such articles, patents, patent applications, or other documents.

The inventions illustratively described herein may properly be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms “comprising”, “including,” “containing,” etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

Other embodiments are within the following claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.
<160> NUMBER OF SEQ ID NOS: 370

<210> SEQ ID NO 1
<211> LENGTH: 1260
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 1

ccatacccg tgttttgggc tagcaggagga aatccacct gcagatgttt gaaccaatgg 60
aacattacca tgaacggcggt attaaagcta tcggctctcg gcggcgccgc gtaaatgtctg 120
tgaaacactg tggcgccgag cgccatgaag gtgttgaatt tttcgcggta aataccgatg 180
cacaaacgcg cggtaaaca gcggcggac agacgattca aatcgtgagt gcgattacca 240
aaggaatggg gcggcgtgcgt aatcagagaagtg ggccggcgaa tgcggctgtc gagagatcgcg 300
atcgattgcg tcgcggcgcgg gaaagtgcgac atcggcttttt tcagggtctcg gtaagggtggt 360
gtaggtaccg tacaggtgca gacacagtcg tcgtgagacgt ggccaagagat tggggtatcc 420
tgcagctggt tcgctgcact aagcctttca acttgggaag caagaagcgt gtagcattcg 480
cggagcagg gcctacaagtg atgtggactc ttcgtcacta atcgcgaacg 540
acaaacctgt gcagctttcg gcggcgcgta ttcctcgcgt gcgtgcggttt gcggcgcgca 600
acagatgtact gaaagggctg tgcgaagatga ttcggaatgt gtttggtt 660
tgaagctgga cttggcgcag tgaacgcacgc taagtctgta gatggctac gcagatggg 720
gttctggtgt gcggcagcttg gaaagctaag gcggactaag gctggtgtgc gcagatgctt 780
cctcgctgt gcgaagatac gcacgtgtcg gcgcgcggtg cgcgtgcttg gcagataacgg 840
cggcgttacct gcggcttttg gcctgggtgc ttcacgcaac gcgcgcggcc gcgcggcgca 900
cgccgcaac gcgcgcgtcc gcggctttgta ctgccttgta cgccggtatgc gtaagcgcg 960
tgcgcgtac gcggctttgac gcaggttatgc gcggactacgc agcggggcttgc gcggctttg 1020
tgaccaaaaa ggacggcttg ccggctttgta cgggttagcgc ggcggggaga gggctggtgc 1080
cgcgcagcc gcggcgcgga cgggctttgta aaagctggttt gcggctaaag gcgcggcact 1140
cgcggcaggg gccgtaggagtc cctgctgctt gcggctaaag gcggcaggtt ctgccttg 1200
cgcgcaggg gcgcggcaggtt cctgccgttgc gcggctaaag gcggcaggtt ctgccttg 1260

<210> SEQ ID NO 2
<211> LENGTH: 1260
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 2

gtaatgcgctc cggggtcgtgaa tgcgctgtgaa ggacgactcaca ttctgcagat 60
gtaaacccta tgtgtaaata cccatagtgc ggtgattaaa gcgactcgc ggggcggc 120
ggcgcgtttcg cggcgcgcgc gcgggtggct ggacgttttg gatctggctc gcgatcgtc 180
gtgatattc catgtgctggt cgcgtggatgtc gggctggtttgc gcggatagtc ttcctggc 240
tgcgctgtc gcggcgcgttc gcggctttgta cgcggtgtggt gcggcgcgttc gcggctttg 300
tgatgtgagt gcgtgggtggc gtcggctgtcc gcggctttgta cgcggtgtggt gcggctttg 360
tgcgctttgta cgcggtgtggt gcggcgcgttc gcggctttgta cgcggtgtggt gcggctttg 420
agatttgggt tctctgacgc ttgtctgctg cactamgcct ttcacctttg aaggcaagas 480

ggttatgca tctcgagcag aqggataca tcaactgtgc aaccgagttg aototctgat 540

cactatcccg aacgcacacc ttgctgaagt ttggggcgcg ggtatcctcc tggctggatc 600
gtgggycgc gcgaacgcgct taacgttacg cgttgccgca ggtacgtctg aactgttatc 660
tgctccgggt ttgatgacgc ttgacctttgc aacagctacgc acctgataatg ctgagatggg 720

tcagcaatct atgggttctg gctgtggcag ccttcgaagac cttgagacag aagttgctga 780

aatgttatac ttctctcgct tgttcggaaga tattgacctg ttgccggcgc gccggctgtcet 840

ggtaaacctc aacggccgcac tccacccctg tcaactgctg tggatagcg taggtaacc 900

cacccggtca ttttgcctcg acaacgacag tctggttataagtgactcctcttgaccacgga 960

tatgcacgc ggcgtcgccct taaaagtttgt tgtcgacgctg atcgccatcg aacaagctcc 1020
tggaattct cttgtgacca ataagcggt tcgcagccca gtgagggatt gtacaccgca 1080

gcacgggat gcgggctcga cccggagcgc gacgccgctt gtaaactgc gtcgtgacsac 1140

tgctgcagaac aacgcagtaaatcaggattcctggtatctcgcttccccagcggtacggaga 1200

agcttgtatgt taacctcaggt gatcggcggg taccagcctc gaattcgttaa ctgatggtcat 1260

<210> SEQ ID NO 3
<211> LENGTH: 2544
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 3

aagcctgca tggcgggctc ctgcttcgc gctgatagtg cccgtagtat cattagccat 60

cttcgaagct acatcatttc cttttttcct aacacgcca cggacatcgc tgggctggcc 120

ctccggtcct ttttatatata ccctcccagag atggagtgca tcgcttccccaa caccatctgg 180

acccgcacttg cgagatact gcgggctgtt ttcgatgact gtcgatgtgtcg 240

tggctctcg ccggtattgtaa acaagggat ccctctctgtg cgccgagaga gattagccag 300

acgcagccgc gacacggcata cttctctgag gcggcgtgc ctgttgcag 360

caggtgtctgc gctgtactgt gacacgcttt gccatcctgg gcgtggtcgc 420

cgctggtttct atgctgctgc gtcgagcagag tcaactctg ttgcgaagt 480

cagcctgaa tagcctgctt cccctgccc cgcgttcgct ggttagtcag atttgctccaa acaggtcctg 540

gaatgctgag ggtggggcct cctgcgcgg aacgaacccc gatttgcgca atattgacgg 600

cccattcagc cttcagtggc ctgagcgcg aacgacgcaaa ttaaaaccct gttgattacca 660

tttggagcgc acggtagctg aacatattct cttgcggaag acacgacaaat 720

atacagcgc gcggagcaaa atttctgctc cctgattttc acctcccctc gacgcgaat 780

ggtggaggtg agatattata cttttctttc ccgggcttcg ttgataaaaa aatagagata 840

acggctcgcc tccactgcccc cccagagatgc gccattaaac acgtccctcg 900

cggcagggga atttcgcttt cgctcagcat actcttccat cccgcgctctc ctcgaga 960

aacacatg cttcatgcttc ggctccagct gctctttttc gcgttttttc 1020

gtaacccca cccgtagact gcgttataaa aacacagctgc tcacacagcgc ggaacccaacg 1080

cgtgcacaaaa acgcgtaaca aatgtgtgta ttaactcgcc gagaagatctc atattgatta 1140

tttgaaagccc gtcctatctt gctatgcatt atcctttttaa cattagcgcg 1200
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TACCTGAGC TTTTATCGC AACTTCTAC TGGTTTCTCA TACCCGTTTT TTGGGTGAG</td>
<td>1260</td>
</tr>
<tr>
<td>CAGGGGAAT TCACCTCGA GTAAGTTGAA CCAATGGAA CCTACATAT GGGGTTGATT</td>
<td>1320</td>
</tr>
<tr>
<td>AAATCTACAG GGGCTGCGGG CGGGCCGGTT AATGGCTTGG AACACATGGT GCCTGGAGCCG</td>
<td>1380</td>
</tr>
<tr>
<td>ATGGAAAGTT TGTAATCTTT CAGGTTGAA AAGCTCTGGC AAGCTGCTGA TAAATCCCTG</td>
<td>1440</td>
</tr>
<tr>
<td>GTGGGACAAG GATTGAAAT CTGATGGGCT ATCCACCAAG GAATGGGCGC GCGGCTTAT</td>
<td>1500</td>
</tr>
<tr>
<td>CCGGAAGTGG GCCTGAGTGC GCTGTTGGAG GTGAAGTGGT CATTGCTGTC GGGGCTGGAA</td>
<td>1560</td>
</tr>
<tr>
<td>GTGGCAGACA TGTGTTTAT TGGTTTGCTG ATTGGTGTG AATGGCTTAC AAGTGCAGCA</td>
<td>1620</td>
</tr>
<tr>
<td>CCAATCTGTCG CTGAATGGCA AATTATGGT GGTATCGTGA CCTTCTTGTG CTGGACTAAG</td>
<td>1680</td>
</tr>
<tr>
<td>TTCCCTGACT GGAAGGGCA GAGGCGTAAG GCTACGCGG AGAGGGGCT AGCATTCAAG</td>
<td>1740</td>
</tr>
<tr>
<td>TCCACGACTG TGATCGATCT GATTACATC CGAAAGAGCA AACTCTGAA AGTTCTTGGC</td>
<td>1800</td>
</tr>
<tr>
<td>CGCTGGTATT GCCTGGTTGG GCACGCAAG AGTACTGAG AGAGGCTGTT</td>
<td>1860</td>
</tr>
<tr>
<td>CCACTGCTG CATGACTTGC TACCTCGCTCC GGTGTGATGA AGTCTGACTCT TTCACAGCAG</td>
<td>1920</td>
</tr>
<tr>
<td>CGCCTGAGCA GTGCTGAGAT GGGTACAGCA ACGTGGTGTT CTGGGCTGCG GAGGCTGGA</td>
<td>1980</td>
</tr>
<tr>
<td>GCACGGCGAG AAGAATGGCT ACCAATTCCT CTAGTGTGGTA AGATATCGAC</td>
<td>2040</td>
</tr>
<tr>
<td>CGTCTGCGGA GCCTGGGCA GCTTCTGTTA ACCTGCGGCA TGGTTGACTG GCCTGTTGAT</td>
<td>2100</td>
</tr>
<tr>
<td>GAGTCGAAAG CGTTAGTAAA AAGACAAGCAG GTACTTCTGG CATTGCTGCTT</td>
<td>2160</td>
</tr>
<tr>
<td>ATCCGACTT CTGGATACC GGTATGGAAT TGACAGGTG CTGCTATGCGT TGTTGGACGA</td>
<td>2220</td>
</tr>
<tr>
<td>GCATGGCGCA GTTGGCAATCT AATGTTGTA GGAATAGCCCA CAATCAGCGG ATGTCACGAC</td>
<td>2280</td>
</tr>
<tr>
<td>CCAGATCGCA GCCATCGCGG ATCAGCTCCG TACCAAGATC GGATGGCGCA AAGCTCGGCG</td>
<td>2340</td>
</tr>
<tr>
<td>CGTCTGTAAG CACGTGCTAG GCTTCTGAGT TCATCGCGGA GAAAGCGCG</td>
<td>2400</td>
</tr>
<tr>
<td>ATCCCGATCC TGCTTCCGGA CACGTCGAT TAACTACTCA GGCGTTCGTT AAATGGTGC</td>
<td>2460</td>
</tr>
<tr>
<td>AACCGTGCGG GAAGTCCCTA TACTTCCAG AGATAGGAA TTCCCAAGA CCAGTACCA</td>
<td>2520</td>
</tr>
<tr>
<td>CTTGCGACAA GGGCGGCGCA ATCG</td>
<td>2544</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 4
<211> LENGTH: 3350
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 4

AAGCTGAGC TGGCTGCTG CGACTCCA AGCTTCTAC TGGTTTCTCA TACCCGTTTT TTGGGTGAG 60
GGATGAACT CAAACTGCGC GCTTCTTGG GATGTCGCGC GCCAATAGCCA CCGAGAGATG 120
GTTCAATCA CTAACCGCA GCACGGCAGA TTTAATTAAT GGCTGCGTGG GATGTCGCGC 180
GAGAATCAGA TGCTATCGAC CTTTCTTGG GGTATCTGGGA TTTAATTAAT GGCTGCGTGG 240
AGCTGCTGCG CGAATTTACTA CCTGACACAG GACAGCAGCT CGGAAATTAAC 300
TTTATGCCAC GCCAAAGAGGC TTTCTTGGGT AGCCGCGCAGC CGGTATACCTT GCTTCTTGG 360
CAACGCTTCTG CTGTCGCTGG CCGTACAAGA ATCCTCAGG TGTCGCTGG GATGTCGCGC 420
CATACCAAGA TCGACACGA CATCCCTTCTA AAGACAGCGA AGTACTCAGT CTTAATTAAT 480
ACCTCGAC TCAACCGCA GCCAATAGCCA CCGAGAGATG TTTAATTAAT GGCTGCGTGG 540
GGCCAGTGG GTGTCGCTGG CGTTCGCTGG CCGTACAAGA ATCCTCAGG TGTCGCTGG 600
ATTACGCTT CGAGCTTGCG GCCAATAGCCA CCGAGAGATG TTTAATTAAT GGCTGCGTGG 660
ttgctccgga caacggaact gtcggtatcgc gtaacctctct tggcggcagg ctagatgacg 3000
agctgccggt aacgctggtt ggcacaggtta tcggcatagg caaaatgtca aaaaatcctc 3060
tggtaccaaa taagcagttc cagcagccag tcgaggtatcgc taacagcacg caggggatgg 3120
tcgcggtgcgc ccaggagccag aagcggcttg ctaaagctgt gattagcaat gcggccgaaa 3180
tcgagaaaga cgcgctactc tcgagatccct cagcactcact gctgtaagcag cgtgattaat 3240
aatctagagcg tagttcagact tgtgcatctt ttcgaggaag ttcctagact ctcttagagaa 3300
taggaactc ccaagcggcata cactaactca gacaaagcagc aagcatccttgc 3350

<210> SEQ ID NO: 5
<211> LENGTH: 2280
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 5

aagcctcgtat tgcggccgctt cagcttcgcct tcgatactgt ctcttaataaa gtgaaatctgt 60
atgtcttttg tgtgctcaata ataactttgta gaaaccagaata atctgttaata caatactcact 120
gtgcggcttta cccgattctgc gattttctaa tccaggctgttg acctctctac cctctcagcag 180
cgtgcggtgcgc aagcctgcttata gcgttgtaacc tcgagagcgg gcgtagtaagcag taacactcttgc 240
gggtgataca tcatctct ttccctttaa tcgctgtgta cgcctgcgg gaagcatctca tctcctagac 300
tggaaccttta tttttttgcc gttttctcttt cactagggcg ccggcggcaagtctggtct 360
tgacccacacta gacacagccat cattcactct caggttgattc ttcatcagct cccgtgtaagct 420
aagcgcgtaga cattcactcact ttcgggcagcc gcggtactgaa gcgtgtaact ccctggcctactc 480
cgggttaatg tcctcttggt tttcatcagc catcactcact caagttgata attgtcagcag 540
gtggcactgta cccagcagta ttcctagacgcc gcctctgccg ccgcagggctt ccgttagct 600
tcctcagactgt ccggccaggtt cttttttgcctg atcactttgc agttgcacagtc cattcactct 660
tgcgtgacgcc ccaggtactg gttatcagec cccactcctgg gcgtccgggc acgctctcttgcag 720
tacgctgcttcat cctcagctct aacctctctag gcggtgactg cggactcagt cctctgtta cattcactct 780
gattcataacc cactactctc taactctctct cggagcatacgc gttctgtaag cgcctcactctg 840
tgagggacta acaccaagact ctcttccgca ctcacccata ctctcagctgcc gggcttactaa 900
ccctcaggcact cctttcagcct tagaaatatt cctctcagccta tttttgcgtt cctttgccggt 960
cgcgaattact cggccagcgc ataataacg caagagaact tattgataac ctaacttttg cttccttccttg 1020
attcattacat tggatacctt caatggggcg ttcatgcaca agcctctctc gccgcagctg 1080
ggccggtaatct gcgggtggac acaagtgctcgc gcggcgtgata cggggttatggtc 1140
gtaaatcagc atcggctacgc gcgtccggctt gcgatttgta cttttttgggt 1200
agctgataac ggtattgccg gcggagcttgc gtaatactcc aaggggctgg ggctcggactg 1260
agtggtgccg cgggcaactt cctggtgctgg gcggcggcgtg cggcggcgtg 1320
ccggggtgtg gcgccttgcc gcggtgactg gcggcggcgtg gcggcggcgtg 1380
gatttggtgata taactactctg gctgctgtaat gactactcctttcacttt 1440
cagctgtctct cccgtgtat cggcgtgactg gcggcggcgtg gcggcggcgtg 1500
acgtaggact actagagcact gcctgtggta gcgcctgaagc caggagacttctgg 1560
ttggccgctt ccggcgtgact actagagcact gcgtctggcttgcgcggtgggcg 1620
-continued

cgccccgggt tggtagaacgt ggaactttgcc gacgtacgca cccgtaatgtc tgaagatgggc 1680
taagcaagta tgggtggcag cgggagagc gtcaggaacc gtcaggaaga aacgtgctaa 1740
atggtctctt ccctccgcgg gctggaagat atcgaacctg tggcggcgcc ggcgggtgctg 1800
gtaaactca cgggagccgt cggcagtcgt tgaagaaacc gtaaacaacc 1860
atcgtgcat ttgctcgcga caacgcgact gttggtatctg taciactctca tgaccccggtat 1920
atgaagagc aacgtgagctg aacggttggt ggcagacgta cgccagctgg caaagctctc 1980
gaaactactc tggtaaccaaa taagcaggtt cagcagcgcag tggagatcgc ctacaaggacg 2040
catgagctgtc cctgcgtcac ccagcgcagc aacgagcttg ctaagtagcgc gaataccatc 2100
gccggcggag aacccggttc ctcgatattc gagatattcc gctgaagcag 2160
gctggttaact aacgtaagag cttcaccaat tggccacact tgcaggggaag ttcctatcttctc 2220
cctagaagatt gggagacgcc tataactcct gacacagcag taaaagggc agacatctgg 2280

<210> SEQ ID NO: 6
<211> LENGTH: 4728
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 6

tcggcgcgttt gcctgatgcg ggtgaaaccc tcttagacctgc gcagctccgg gacaggttcag 60
cagctcgtct ggaggggaca ccaaaagggcg ccaagccgct gcgggctgacgcc 120
ttgccgctggt cttggcgcgg tcttcacatt gcaggtaaga gcaggtaagg taaataacgc 180
acatgtcgc ccgctgaaat ccagccgagat ggggtgagag aatagccagc atcaggggcggc 240
attgcccaatt ccggctgggag aacgcttgagc gcgggcggat ggtggcggcgg ttccggatct 300
tcggcgctt gggagagcgg gggagcgtgct ccaagggctt aacctgcttga ggggcgggtta 360
ctctccgatt cggcgtgctt cggcgtgcct ttcattgag aaaccagag ccagcggcgg cggcgctgctg 420
aattgatag cagcagttgg cctggcggcg tccggccgcct cggccgctc acacccgaaa 480
aatggctatc accgtcatgg tccaggtctc tgggccggcct gctgctgctg ccaaaacagtga 540
gctgagagct atttgccgca tcacacactt gcagctgctgc aggatagcct 600
tctgggtccg ccggcgtggt cttggcggag aacgctggag cggctggagc gagctcaac 660
aaaaatcgc cgcagccaggg gcacccgttct cggctggagc cggcgccgggt tggccgactt 720
ccagcagcgt tcggcgcgg ggtgtaacgc aagtaacttc gggtgattcaagg 780
gccctccctc caactgcgag cccaccaact cggctccatgc tggcagcaggg cggcctgagct 840
gctgctcttc atggctcagc ttcggcagaa cctggcgcgg cttggccgcgc cccgcggctac 900
tcggcgccga tggccggctg cctggccggg ccttggctttt cggagacac gcggcgcgctc 960
caagatccat ccctcagcgc ccccggcagt aaataatatt ctgcaacaaa agatcgtttaa 1020
cggagcggta ggagtgttta tccgcatgac gatgtgaaaa gagatcggcgc cgggttactgg 1080
gtagagcggc atcgggttgt gcaccccccg cccagccgca accagcctcc 1140
aaaaatcgc tggcccgact cggctggcagc cttggttcttt gcagactgctc 1200
cctgctgttct gcggcggcagc aaactcatt cggagacgtt ctcagcggcag 1260
cctgcggtgc ccagagacg agggttaggt tgcagctgctgc gcggagcggc 1320
gattccgc cctgcggatg aagcagcta tggcaaatgt cggctgaacag 1380
aactctcggc gaaaaagctg cgaaaaagtgg ttactgtcgc tgaatccaca gcgaatagcgg 1440
atgctgaaag cggtggcctg gcttggtgcgct gagaagatgc ggggtttccaatcg cagctgcaag 1500
cgggtcagg atcgtcggag cgctcgatccc gtgttgctgtc tagatgcgcct attgagctta 1560
cgctgcgtaa gagaattactg atcgcacagcc gcctgcccaatc tcacccacat gcgcaaatgg 1620
tcctccagcc aagccagaga caatgtgaga cgtgatggcgc tggttctccag gttctctgcg 1680
aactgctgttt taacgacgca gacagctaat cgatcataac acgatctgacg actgctgcggc 1740
gaggtaaat ctattttcccct ttccctgtctg tcactctgtg caaccagctg tgcacactgc 1800
tgcaatacgc ctgctgtaac gcgcagatgta gaagcatact gcccatacag ctccttgtgc 1860
agcactgctgg tccgcacgcc gagaattgca aatcgtccag gcggagccatga cagcaacttg 1920
gtcgacacac gattatcgtgg atgtctcatac agatgcccag catgatccgc tacgaacatgg 1980
acgggtgccg aaggtgagtt atagggcctg ccattaacaca catgatatac cgtgcctatgt 2040
tgcgacactca caattcccagc aaaaatctga tgcagttcag gcacatccgct ctgccccgccg 2100
cggggttcctg tcggcagcagc cgctgtaacgc gcggagaaaa aataccacat actatactgc 2160
gtcatcagtc cccccctgcag tgcgtcaccac gccgaatagtgta taatcagcag gtcaggttctt 2220
taatcttaaat ttgagcaaggt aaaaaagttcg aatitctcaag atacagctgg 2280
aatttccaggg aacttcgggct agcatcatac ccacaccagag cagcaatcctg tcaatcactt 2340
cagctcttcct ttccctccttgc tccaccatcag gtaacatcag caggtgatgg 2400
gatctcaggc gctttttctgg ctgctgttcga tgcattccag cagctccacct tgcgtcctgt 2460
cagctctgag ggacgcgcgggt tcggcagcagc gaaatcgtgc ataactgtgta tagctgcttc 2520
cgagtgtgaaat cgggtactcag cctacaaccct cacacacat gccgaacggcag caacataagtt 2580
gtatcagcgg gggtgtcctaa tcaggtgactg atcctctcgag aatctgctttg gctggtctccg 2640
cgctttttcct gccggggctcct gttggctgctg gccgctacaat agatcgcagcg cagccgctggc 2700
gagggagcctgtggattgtgggtctcctt cgcctgttcct gtcacatgcag ggcctgcggtc 2760
cggctcggctgt ggagctgacgc gcgggttctagctgcctgg ttcgtcagcg 2820
cgacaggg gggagctgta cggacatgg ccggagacata tggagcagaa cggccggagga 2880
acgggttggc gcccgtggcttt ctctagctctc tggccgctcc ctgagggtaag 2940
acaaaaatagc acgcatcgagc cacaggtggcc gcaccccgac acggactataa agataccacg 3000
cggattttctg tggagtcctc tgggttgtcc gcggcggagtt ccctaggtctg cttaagctgat 3060
acgtgtgagcc ctctttctccttg agggttcgcttc ctactgtcctg aactcttgaggt 3120
atctccagcgc gctggctggtc gttgtcattc acgggcctgg cggagtgaaccg 3180
acgcgcgagc cgctgcttcagc ttcgccttgatg atcgcatctgc tgccaaacgcc gtaagacagc 3240
acctactgcc acggcgcctc ggcatctgtgta acaaggattag cagagcggag tattgagcgg 3300
gtcctgaca gttctggtgag tgggtgtcctcg ctaagggctt catcagatgctt acagtgattg 3360
gtactggcgc gctctggtgag ccaggttactct tgggaagctg tggagtgacttg ttctgtcacc 3420
gaaacctaca cagcctcggt aacgctgggt tttttgtgct ccagcgacag attagcgcgca 3480
gaaaaagacc atccagagct ctgctctctca ggggtgctgcggc gctcagctgg aa 3540
agaaaaaacg caggttaaggg attttgtgctg gtagattct taaaagagaatg tttaaccagta 3600
tcccttttct taaaaatgaag tggggttataa caattataag tataatgag taaacctgtt 3660
ctgacagtta ccmatgattt atcagtgagg caactctcgc aagcatactt cttatctcgt 3720
catcataagt tgcctgaacct ccgctgagtt gaaaaacta gataaaact aatggagag 3780
cagcaccgg cagcaccgg gcaagtgagt tcgtctacgt ttaaatgcgt 3840
cgatcagtc ctatattctg tgcggggaag ctagaagtta taagtaagct caggataaatg 3900
tgcgaagtct tgtgctgtct gtcacagga ctcggggtgt ctgtgagttc 4020
catttctcag ctctggttcc caagatacca ggccaggtta aaatgactccc esgttctgca 4080
aaataaagcgt tgcctccttc gttcctcctg cttctgctcg gaatgatgtg gcccacgcagt 4140
tatactcctt ggttatcggca caatcgtcga attctctact ttcgctagcct cggctagata 4200
tcgaatgtgt gcctgtgtag tcttcaacca aacgtcattcg aagataagtt atccggcgtcc 4260
cagtttgtc tgtgctggtcg tcaatacggc ataatcagag gcaacacatc aagacacttt 4320
aatgtctcat catgtaaaaa gcgtcttccttg ggcacaaact cctcagggatct ttcgctgctt 4380
tgcagatcct tcgatctgat caaatctcttg acaaatctctt caaatctctct ttccttcttt 4440
tcaacagctg ttctctgctg gcaaaaacag gaagcggaaa gcccggcaga aagggatatt 4500
ccggcaacg gaattgtgtg ataatataaac cttctccttt ctaatatatc tgaagctatt 4560
atcagggtta tgttctcag aacggtataca atattgaata ttatctgaaa aataaacaata 4620
tagagggttcc gcgacaattt cccgtaaaac tgtcaagctga cgtotaagaa accattata 4680
tcgtgacatt accatataaa aatagcgtata tcagctggcc cttctgc 4728

<210> SEQ ID NO 7
<211> LENGTH: 1440
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 7

gaaactgggc gtttcttaag cttcgctgaa tggactaca aatagcagat 60
gctgtcttgg gaaaoaaggg cttcgctgaa gatgattact gcgaagggcg ctgctgctg 120
tttacgggt gatcagactg cctgcgctcg tggattactt actagataac ccgaaaca 180	tacatttttg cttgtcctgg gtatcattg ggtacgtaa gaaatcctt gcaatccttcct 240
tgccagggc agctggcggt aatcttctct ttcagctaac ctgacaaaacc ctttctctttt 300
cactcactc gogtcgacaaaa ccaoaacgg ccaoaacgac taatccttcct gcgaatcctt 360
atcagacaa ctcgcgctca accgactcct cacaacatcc gaaatcctt gcaatccttcct 420
tcgcgctca ccttcgagc ggttcttcgt aatccttcct gcgaatcctt gcaatccttcct 480
taacgtttct aacgctacat gttgagagct caagcagcct gcagcttgctg cctagtcttc 540
tcgcgctca gttgacgtagc ctcgcgctca ctttcgagc ggttcttcct gcgaatcctt 600
ggaagcgcgg aagggcgttt ttcagctaca gtcgctgctg gctgctgctg 660
tgcggggcc gcaaaaccct ttgctgtaac ggcatcagct gcgggattgc aaatggggaa 720
gctgcctag gctgctgcct gccgtgcggt ttcgctggtta actaatctcg gaatgatgtg 780
tccgctgac gggctgacaa ccgaaaccct ccgctggtta ctttcgagc ggttcttcct 840
gcgcagggc gcggccgac ccttacgtag ctttcgagc ggttcttcct gcgaatcctt 900
cacgtgtcc gggcaggtgg aagaagggag gtctagtagt gacaacattt gcgaatcctt 960
<210> SEQ ID NO 8
<211> LENGTH: 1560
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 8

gaatccagcg gttttttgaa ctaggtgttaa tgaattcag caggatcaca ttctgcaag  60
gtcaagcgc agactttatag tctgttttgt ttatatatat aagttagttg tcaatggaga 120
aataaagtg aaacaacaag ctatgcaet ggaacactta cgttactgtg ttacccctgt 180
gacaaacgcc cggacacgg aatgctgtg tctggaacac cggacactgtc agggcagat 240
tactgcaccc ggggtctggtg ggggttttaac gggtgtaacag attggctgct tttggttctc 300
tcttcttggg aaactcctgaa aataattttg ttggtgatt ggggtgaag gggaaggactc 360
ggaattaact gccaagctga attatgcgca aagtgccggc ggtttttttta aaggttataa 420
tgcctttcag cttaaacgggc aatactctca ctatgctgct aataaaaaaa cggcaaaccc 480
gattgaacg acagactcgg gctgatcagc aacocggcgtg tcaacccgtg tcaaaaocct 540
tacggcgcgc ctgggtgcgcg atataccag aaaaatgcac ccaagacttc tggaaatggcg 600
aaagaagcga gttcgtagga cccgtaacgt tcctacgca gatggacagg atgcaacgc 660
cgtgctgcag tgcgtcgcaag tgaatcctcg caatgtctac gttcgacagcg cgcacagttg 720
aattgtcgg ggtacacgtc tggaaaaagg ccggaaagga tcggatttcc aacagcgtatg 780
taagcgtgct gccgaggtta cggtttggcg cggccgaaac cccctggcgct gaaagcgcc 840
cgctggttag tggcgggaag aacocggcgc tggacagcga cgggccgctct gttatactgtg 900
gtgtagcagc gttcttcagc tgaacctgcg gcggagagcg aatcagcaaa aacocctgctg 960
tggtcatgta gttcgcctag catgtgctga agccggagca aatatgtgcttgctctgtctt 1020
tggcataata cttgacaacc gcagctacctg tccgctcaata cggccacgta atgatcggt 1080
accacccctg ggcgcaagta cggcaaaccc cattgaattg ttggaataaa aatggaaaaaag 1140
cctttttctgg caatgtaagag tgcctcatac gataaaccag gatcatgctcg cgaatccttg 1200
tggccaaat ggccgacagg tggattcgcga tgaagcgctga caaaccggcg tgggatcgc 1260
aataaaggg ggtatcagcg tgggtcatacg ccggctgtac ccagcccacc ccaagcgag 1320
tggtgcgcgg gataatacgcg caaaccgggcaattagcccga aataatgccg 1380
agtttgcgatg tggggtcagc ggaactccga agagggacct caagaacact cggccactca 1440
gttggtgtaatt ggccgtagtg gccgcaagtg cgccaaatgt gttggactga cgcacgcagc 1500
cgatottttc ttacccatga aacocgcttt gggtgtagaa taatctagag gatocccgggg 1560
<210> SEQ ID NO 9
<211> LENGTH: 1500
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> Feature:
<222> OTHER INFORMATION: Synthetic chimeric nucleic acid sequence

<400> SEQUENCE: 9

gaatcaggc gtttttttaga cttgctgttaa tgaatttcag caggtatcaca ttctgccgat 60
 gaacctttgggg aatcgacgtg ttatctcgat acggtcttta ctctcctctg cagtattact 120
 gctcatgctt gttctggaaa acgggtggtgc tcagggcagat attactcgac ccgccggtgc 180
 tcgccgttta aacggggtgta acagtcgccc tcctgcgtgct tctctttagag ataaacctgct 240
 aaaaaatatt attttgctga ttggcgtatgg gatgggggac tcagaatata ctgcccagcag 300
taatattgac gcagtcggtgcc gcgcgttttt taaaggtata gatgcctttac cgcttacgag 360
gcatacact cactatgcgc tgaataaaaa acgccggcata cgggactacg tcaccgaccc 420
gcgtccataa gcacacgctgt gcacacgggtc tgtcacaacc tataacggcg cgttgccggtt 480
cgatattcag gaaaaagtc accaaccagat tcgtgaaattg gcaaaaaacgc caggtctgcg 540
gaacacgtac gtttctacag cagagtgtcga ggatgcacaag ccgggtcgcgc tggggtcaca 600
tgtgaaccttg cggaaactgtg acggttccag ccggacactgt gaaaaatgctc ccggtaacgc 660
tctggaaaaa ggccgcaaat gatcgatttc cgaacagctg ttaacccgct gcggcagagt 720
tacgctgagg gcgcgcctttc aacccctttgc tgaacgcgcgc accgctgttg aatggccacgg 780
 aaaaaaaggc gtgatgacgc cacagggccg tgtttaacag tgttgcgagct atggtctctc 840
aatgatctgg tgaacggaag gaataacgca aaaaacctct gttggtccttg ttttgacgag 900
catagctgaa gcgcgtcgcg tggagccgaa acgcagcttc catgggaatgt ctgataagc 960
cgggacgtac ccgtgccgaat atcggcagag taattcgcgtg tgcacaccctc gcggccgagat 1020
ggcgacaaaa gcgttccagat gttgctgttaa aatagagaag gcgttccattt tcgcaagttga 1080
aggtgctgattc atcgataaac aagttatagc tggctaatct tgtggccaaa tgtggcagac 1140
ggtcgatttc gatgacgccc tacaacgggc gctggaattc gctaaaaaggg aggtaaacac 1200
gcgtggtcata gtcacgccct atcaagcggca gctcggcagat attgttcgcc cggataacaa 1260
agtcgctcgg ccctcccaggg cgtcataaatc caaagagccc gcagtctgtag tgtaggttag 1320
cgggaacttc gaaggagag tacaagcaaca taacggcctg cagtggccgtata tggcggcagta 1380
tggcccggcc gcggcactgt gtttgctgact gcgcgcaccc accgtacttt tctatcccatac 1440
gaagcgcggt ctggggtcga aataataac ttagggatct cccgggtacgc agctgcaatt 1500

<210> SEQ ID NO 10
<211> LENGTH: 3968
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 10

tcgccgttta gcggtgatagc gcgtgaaaaa gctgacacat cagctcctcg gcagcggtcag 60
 caggtttgct ctaagcggat gcgggccgag tgcagccgctc tcaggcgcgct tcagcgggtg 120
 tgtgggcggct cttgggggggg cttaactcag gcgcgtcgcgc gcagattgta ctgagacag 180
 accatatgcc gttgaaatct cccgccagat gccctaggag aaaaaaacgc ctacccggc 240
atcgcatt caggtgcgcc aacgtttgag aagggcgatc gctgccccagct tttcgtatat 300
tacgcagct ggccagaagg gattggtcct caggcaagtt aagtgggtta aogcagcggt 360
tttcccagtc acgacgtgtg aaaaacgaag ccagttgccaa gttgccagc gctgctcttt 420
tgcgtgctg cttcgagctg gtaaggcga caaattctca cggcttcgag tggcgaggtgta 490
cagagctcata attaaacac taaacagcgc gtaaacaacc gtctgctgttg tttacgagga 540
taaacatct aacgttctcg cttaaaactc gggagccaaa ttaagctgtat ttttaacca 600
atatcagatg gaagaagata cgaatagagt gtgcctgctg acatctccac aagcaatgata 660
ggcgcgctcaac cgcaacagcg gttaaaactg tcttcagctaatctataagt gtttaagat 720
ggataacag cagacatgaac cacaacccag acgcaccagcg atcttcacagt ctgcctcaca 780
atatattcga aaaaagtatt ttcacagctgc ctctgtgaaa agcatctgtga aatctttgc 840
cattgtcgc ggacgctctt atcaccattt ccatttccaa gaaaccttga tttcggctat 900	taacctacg gacccagaga ggccggtcgg cgcttccgac ggaccgattg aagggattca 960
atttcgctgc atatagttcg atccttcacat ctttctcacc cggaaacgct tttgacaacg 1020
ggctgctgtt gttgaaattt cggcgggaag gttgtgtaac ccacaggctg cggccatctg 1080	taaaaaaatg cattatgac tcaaggtatt ttggtgccag ggaggctgtg atgcccagaa 1140
aaaagcgagg atacaagccag atcccaacag ggcagatgac tcaagccatgt tgcgtgaatt 1200
gactaaggt gcattggccc agatagacac gcggacgccg ctctgtgcaag cgttgtggttc 1260	atcgggatttt gctcagatca gatctgccag ttcacgctatt ctctatttattt cacatgctca 1320
cggcggggtg tggtaaatg tggcgcgctt tctattctgta aatataag attttctttg gcaaaaaatta 1390	tcttcgctgg ttctggcaagtt taaaacttttaaatctcgtat gtagctacat gctcagtgaag 1440
atatcgcttc ccaactattg cctggcttct ctttttattg ctttcgctga 1500
cgtgttttta aagatattgc gatacctat ccgctgttg ccgtgtcttttctt 1560
ttgatcgcgg tgcggctgct ttttatcatc atcttttctc cttgccttcct tcaagcattaa 1620
tctgtgtaa cacaagattg gataaatata cagatagpga atttatttattatt atttatatttt 1680
ttaaggagtag ttcctcagctgccgatcctag ctgcctgccg gtaaagcttc 1740
cataattcgg tcttctttg gctctttgaattt cttttgtaggg ctaaacttct 1800
cagacgggga acgtagataa ttaacggctg ggctgcctaa cttgtagatat cctcactcatt 1860
aattgcgcgt gcggctcggc cgcgtttcga gtgggggaaac tctgctgtgc acgtctcattta 1920
atagactgg gcggcgccgggg gtagctgatttc ggctgccctt cggcttcctc 1980
cttcgctgat gcggccggag gctctgctgtc ggtccggcgg ggtccagctgt ctcacattc 2040
gccggatagatt gcgggatttaa ggtccttcgg gaaagcagca ggaagaatca tcgtagc 2100
agcggcagcag aaagcgagta acgtagaaaa ggccgctgtc cttgctgttttt ctcataugg 2160
cctggccttc gcggcgtcctc ccaaaaaaatgc ccgctgctatt cggagggcag cacaagcag 2220
agcgtatagag ccgctgccct cttggaacct ttctgccgctt ctcctgcctc 2280
gccgtcgcg ccctccagat ccgctttgatt gggagcggtg cggccgccgg 2340
tcattactct gctgcgctgtg attctcagtc ggctgttggc ggtcctccga acgctggcgg 2400
tctggccggg acccggctgc cgggggcgcag ctcgctgtttcg ccagacatctt cagggatttg 2460
gtcaaccccg tggagccaga actattacgc acgtcgcgca gcaagttcgcgtaaagagagt 2520
-continued

tttttatcgg tctcgtcag taaacgcgca aacggtgtta tagcgatcgc ggttaatgtac 600
tgtttatcag tcaaattgat gaccocttg g gaaatccca tcatgattgt gggttaaagt 660
ttgacgcgct ccagccacata gttgatgggt aacggtctgat catagttttt ggcaataaaa 720
ccccagatct ctgaaacata aataggcgcga cgctgcgatttt tgggtgctgc 790
gaggttttta tggcccaagaat cggctcccaag ccaggggga ggcaatcagcag 840
ccaatcctct caatgccgag ctggtgcaatt tggcattggactcggttaaattt 900
cagcgcggcc ctccaaacgg gcaaggttgc tgttgtgcgca ctgatttgag acacatgagg 960
tgagtgaaat gttgataacag gtcttttactcc aacggcagg agaaacaag atgcaatcgc 1020
ggagaaaaa tagcagatct ctgagcaggt cccgtagatg ttagtgggag ccgtgtacag 1080
gccgaaaca gcggatatag acctgattg atatcacttt ttctcatgtt gatcaggtat 1140
tccacatgc atcagaaagg cacatcactc tccacgctag catgcagcagt gctgggaggg 1200
atgtatgacc gtgggcaaaac cctaacttcccc acctcgtgttt atccgcaata cagcgcgacg 1260
gggtctgcgg ctgcttttttc gtcgtcgtgc ccaataacaac tccatatatt catggattg 1320
tcctttcttc ggaatatgcga aatagtgcga gacgcggagg aggatgacgc ccagggggag 1380
cacgggggg ggccaggggg cacaggggag gggagggcg cacaccgatc caagggcgaag 1440
aaccttgcgt tcgctgtacc aataactcag ttcgtctcgc ttctcgcacag 1500
acccgggttt atggggaagt ctcaaacgaga aacggaggtc ttctttcgag attcgctgc 1560
atgtgacgat tattcagcag ctctctcgac gacgcgggag gggagggcgt ttctgtcagc 1620
cggttacag gcaattcag gttacatgtt tttcctgtgt gaaaattgta 1680
tccgcaacca attttcacaac aatcagcagc cgaagcataa aataagtaaag cctgggggagg 1740
tctatgagtt gctgaaacta cattttactgc cattgcgtcag cttggctcacatttgggagg 1800
aaacgttctgc tgtcgacggct atttagatag cggcacaagc gcggggaggag gcgggtgtgc 1860
atatgggagg tgtctcgcttg ctctgcaact tgcacggatc gctgtctgag cttcgcctgc 1920
gccgagcggta tccggctcct ttaaagcggt attacgctta tccggcttcct cgggggataa 1980
cggcggaaaag acacatgtgag ctaaagggcc gccaaaagcc aggaggctga aaagggggcg 2040
gtgccgggtg ttttttctca ggtggcggccc ctctgcagag cttccacaatt caaacaggct 2100
aatcagcagc tggccggagc cagaccagact ataaagaatat cgggggtttc ccctggagg 2160
ctccctctcg gcttgctcct gcgcatctgt cccgtgacgt tccgtgctgct ctggttgggg 2220
ccctgggccag aggtgctgcgct ttctctcatag ttctcagctg aatctctctc gttgtggttg 2280
ccctgtcctgc tctcaacctgg gctgctgccgt tccggcggcc acgcgtgcag 2340
ccctgacggt acctctactgc tgtggcccaaa cgggtgaaga caagacttaar gcacactggc 2400
aacgcagcctct gcgtcagagc gcgttatgta gcgggtgctga caggtttcttc 2460
gagaagggcct ggtcaactggt ttagtacagc agagagagtaagaaggagctcttgactacagtctg 2520
gagggcgggt aacaaaagag agaagacttg gctcgcttgta gcccggcaca aaaccaaccg 2580
ttgtctgggt ggtgcggttgg ttggcaagca gcagatttcg cggagaaaaa aagggactca 2640
agaagactctc tgtgtcttct ctaaggggtgc tgcgtcctcg tggacgggaa actcagcgtta 2700
aggatatttc gttcagagat cttcacaaggg gatctctcaca tagctccttt taaattaaata 2760
atgaaattttta aatcactaatc aatgattata taagttacatt ggtgttgtgatactgaaatg 2820
cttactcagt gagggcaacta ttcgaagctat gtgttatattt cgttcaccca tagttgacctg 2880
aactccagctg tgttagaata ataagcaatgg gagaaggttta ccaatctggcc cacagtgcctg 2940
aatggatcagcg caagacccacc gcctcaccggt ctcagattta ttcagaaatat accagccaccg 3000
cggaagggctgg gagagcggcag ggtgtctatc acacttattcacc ccccccttac agtttattaa 3060
tgttggcggag gaaggagagag taaaggtagtgg gcacagtatcct acagggcagaa aagttgtgtgc 3120
catcgctcga gcgggtctcgag tgcggtttct gttggggtgctg atgtgcttcgcccc gagtaacga 3180
ccacggggttg ttcgacagtac ccctcagtgtg tgcacaaaaag cgggttgatcgc 3240
ctcggcttcgcc acagatcttg tggagagatg gtggtggtctgatacagcctg 3360
ggcggcgggtca caatactctgct ttactggtctaat gcctcaggta agaggttttccc atctggcatg 3360
tggagacgcctc aacatggctctt ctcgagaata actgtcgggc cagccagcttg gctttgccc 3420
ggcgtcaata cgggataata cccggacacagt ccacagtccatct ttttaattcgt gtcacattgg 3480
aaagcgcctttc cgggcttcag cagttcttccag ctcggctgag ctctggcctgcc ccgtcga 3540
gcagaccggc tgcgcagcct cagcattctgc gcacatctcctt ttttcacaaa aacagcggtcc 3600
gttggagacgag aaaaagggag aaataagggagta ctgaggggccc aacatgcgagtc 3660
tgtgaatactta atactatccttttaa ttttcgagagc aattatcaggg gattttgtgc 3720
ctctgcggagctc tttctcagatt aagttattttat cgggcaatttt ggcgcgggct cttcgcac 3780
atttggtgacag ctcagatctat ttgctacattactttttc cggcggcaca cggcggcaccg 3840
ttcgagctag gatcggcagct tcagcgctag ttcagatcaga agaggtctattt atacagctta 3800
tcsatatg ggtctttctcg tc 3872

<210> SEQ ID NO 12
<211> LENGTH: 3934
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 12
tcgccggtgttt cgggtgtatgc ggtgaaacc ttctgtacagt gcagctcctt gcagagctgta 60
caggttgcttg tggacggcag gcgggagaga gtaagggcggt tcacggacggct gtcggttttg 120
ttcgggggttg cgctggccttg cttaactatgc ccagaatcgag cgcgtgatgcta ctgaggtgtg 180
acactatggc gcggtgttaatat cggcgcagct gcgtgaaggg aaataacgcacttc acgcggcggct 240
atttcggtctc ctcacgtgctgcc aacttgtggcg aagggccata ggtcgggggcc ttcctgtcatt 300
taggcagctgc ggcgcagagg ggtactgtctgg caaggcgatt aagtgggta acgcggcgtt 360
tttccccagct cgcagcttgg aaaaagcagu ccaagttgct gcctgtcagct gtcattattgc 420
tgtgactgta cccattgtag cacttctgct ccagccacaccct tcctgactcatttcct tcccacg 480
gccagagact gcacgtgcggct ggccgggcgtt gattttttta atacccgcca gaaataagagtt 540
tggagactgca aacccaactgc gcacagccagct gtggcgatct gcacgctcgttc ggtgctacaat 600
gtcgagctgct cgctggcattg acaattgtcgc ctcgctatcctg cccagcctacat 660
ttcgggggttg aagttgagct gcagccagag gcgggacgcag gcacatcgtgtcg cgcagccgttg 720
ggagataata acattctttgt tgcagcgttgc actgactgacccc ttcgctg 780
cyctttactt catggtgatg gagagacgcgt cttcagattt ccctgggccgc cagtaaaact 840
ttcgctcttc gagaagtcgcgc ggtcgttcag ctgtgccgcc cggctggccttc atcgcct 900
atgtgtggct ccaaaaagcattgagcgc ggtggcg ggccttcttgc cgttctcggcc gcggagaccg 960
-continued

ccegtattgg caaatatatga cgccaggtta agccatctac gcagttacgc gcgcggacga 1020
aagtaaacc actcgtgata cattcgcac gcctccctag gacgccgta gtgtgagact 1080
tctctcgcc ggcaacggaa aatatccccg ggtcggcaca aaaaatctcg tctctgattt 1140
ttcacccccc ctctcgcgcc aatgtggtga tttgagatatt aaccccttcat tccagcgggt 1200
cggtcgtatga aaaaaacgag ataacgcttg gccctacatcg gcggtaaccc gcgcacccaga 1260
tgggcatatgg aagactaccc cgccccgaggg gtagctttttt cgccttcgcct cttagacgat 1320
atactccggc cattcgcagga aggaaccatg tgcctctatt gcacacagca ttgtgcgtcag 1380
tggtccttact ttgctgtata ctcgtaaccc aacccgatttt ccgccctttag taaaatctatt 1440
tgttaaaaaa gcgcgaccaag agacccgagaa aacccggttc aaaaagctgt tataataacac 1500
ggccgaaagg ctcacatttaga ttatttgcac gcgcctccac cttgtctatgc catcgactttt 1560
tatacatata gattacggta atcactctga cgcttttttag cgcaacactct taagttttttt 1620
ccataacggct tttttgcccc tagccgaggg aatcaccctc gcgggatgcct ttctagggat 1680
cccgcccccg agcctgcccatt tttgtacgttg ttctctgtgt gtaaattgtt 1740		
tttcostgcata cacttttaaa aacacaccaac gcgggagcac aaaaaggttaa aactctggtgg 1800
ggttaaatt ggatttgagct ccaattgctg acgtggtggc aaggtggggt aagttttggtt 1860
gggacaccatg cgtggtcggc tcacctatggag aatgcgcacaag cgcgggggag agggggttttg 1920
cgtattttgg gcgcctgccgg ttcctgctcg cttgcggcgcg ctgggggggt 1980
cgggagccgcgcccacagcgtcactaata gccctccaga ctcacacattt atttggggtttt 2040
aacgcgcgaa agaacacttgtg aaaaaaacgcc caagcaacagc caggaacgc acacacggtt 2100
gggtgcttgg cgtttctccca taggcttcgc cccctgcagc acgtcacaac aatcgcacgc 2160
tccagctga gcgttgttaga aacaggggca cgcgtatccag cgcgtatcagc taattgaagt 2220
aggtctgctgg ctgcggctctc gcgtctagcg agggaaaaag cctgattctc gcgggtatttt 2280
tctctgtggc agaagtgctgg gctttctcatc agtctcaccg atggcgatatc aagtctggtt 2340
taggtctgctg cgtccagagct gggttctgttg cccagaaccc caagctccgc cgcaggtgc 2400
gcctatcagc tttgtgctgc acctccgattt aacccggttac gcagaacactc taagcgcacgt 2460
gcgaggagca ctgcgcgtgc gattacggct cgcggggtag gcctggtgctga tactaatgcc 2520
ttggagcttg gccgtacactc cggctctcaat agaacagacag tatttgggtat ctgccgtcttg 2580
tgtaagcccc ttcgggttgg aaaaaagaagtt ggtagctttg gcggggccaa aaaaaaccac 2640
gctgtgtatgg tgctgtttgg tgtggcaga cggcgacatt gccggagaaag aaaaaggtct 2700
cagggagagag ggtcaggttttg ttctacgggg agtggcgacga aactacgagt 2760
taggggttt cggagctgag cttataaaaa agatcctctcc octagatcct ttttaaatatt 2820
aaatgaagtt ttaatacata aataaagataa tgtggtctgta cagttaccaaa 2880
tgcctctca acgtgcggccatt cttgcttctg tctggtctttcct catagttgcc 2940
tgctctccgg cgctgtgtgat aactagcag cggaggggtct tccatctgg cccggctgtct 3000
gcaatgacac ggcggagacc gcgtcaccgcgt tcagcgaattt aacacgacgcag 3060
gccggaggg cggcggccgcag aagtgttcct gcacacttat cgcctcgctcc ccagctatt 3120
aatgtgctgc gggagagtag aagaagtagt tgcagggctat atacggttcg caaggtggtg 3180
ggatagtgtct cggcgacgtc gtcgtcggct gtctggtgctg tacgtgcctt acctcgctct 3240
<210> SEQ ID NO 13
<211> LENGTH: 6013
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 13

tagttattaa tagtaaatca ttaacggggtc attagttcat agccoaatat tggagttcgg 60
cgtatccaa cttacgtaa atggcggcgc tttgctgccag ccaccaagac cggccggc 120
gacgcactaa atgacgcatt tccgacagtt aacgcaatatt gccggcgcgat gccgctgctg 180
ttttggcttgc tatattaggt aaactgcaaa cttggtcgtat cttcaatgtgc tcatatgctg 240
aatcgaagcc caactctgac gcactgaagc tctgctgacgc ttagagcggc 300
catgtcgtgat tttttgggct atgacataaa tggagctgga aggtctgctttg aacgaggg 420
atttcaagtt tctcactccaa tggagctgca tttggtttttc cttgagcacc aacaactttccc 480
ggaccttcag aaagtctgta aacactctgac cccacagctg ccaatattgctg ctggtttctg 540
gcctgttgag cttgttataaa cgcacaccag cttgttcgtga ttgactgctgg actggtacct 600
cggtggacc ccatgcttagg caagggcgag aggctgttca cgggggttgtg gccaatactct 660
gtcagcgttg aggagcagtg aacggccac aagctcagctg ttcttgagtc ggccggccg 720
gatgcacaggt gccctctctg tccacagcc ttcagccgca ggtggccccc 780
ccttgccgctt tcctgctcag tccctgcttg ggcagcggcgc gcctgtttgg 840
gacacatag aagcaggctc gtttttccag tccgcctagc gcaagggctc gttgctgttg 900
cgcacatatct tttcgggaga cggcggccag tccagccggc gcggcagcggg ggtcagtctg 960
ggcgcacccc ttggtaaccg ctcgttgctg aacgggctcg acctcaagga gcacgcgc 1020
acccagctag cttacgtcag gtcagcctct cgacgctcg cttgcgctg tcctgctgtg 1080
aacgccgct agcgcagctc gttggagcac gcccggcgc gcgtgctcggc ctggtccggc 1140
agcgcgctc gtcagcgagc ccagcggccc tccgagccac cggcgcccc gctgttcgtg 1200
acccacagct actacagtcg ccccgagcgg gcctgtgcgg aagcagccgc cggagcgcg 1260
gatgcacattg ttgctgttca gcggccggca tcctctgggg ctcgagctgc 1320
ctgtacagt cggagcctcg atctcagact taataaacag cggcaaccttg tctgatgtcgt 1380
taactaatat gaaacttgtc ggtctacatc attcactttt ttttccaaac agggcagcga 1440
cgtgtgctgg cgtgctccgg tgcattttttt aaataccgca gagaatataa gttgatgc 1500
aataaaccgac ttgctgacgca cggtgtgcgtt cagcagtccc agggtgctcc tataaaca 1560
gcgtgtgagc tataatttct ctttgccgca gtaaagcagc tataaatcct aactgtggcg 1620
gaaaagatg cagcagcgcc acggccacca acaaaacctgc tttgagcgcg tggggtactc 1680
aataatttgc tggcaggtgt gatcgtgtat gtaattgacaa gcttctcgta cccgattatc 1740
catcggtgaa tggactgacaact gtttaattgc tttctgtcgcc gcaagtcgaact atctgattc 1800
cagttttatc cggcagcagct ccggtgagcc cctgtcctgc tgcgtggttc taatggttgg 1860
cccaacagcct tcgctgtgatcg cggcgtgcctt gggcagcaagccg gaggagttat 1920
gcgttaattg cggcagcgttg aagctgagtc actatcttgg ccctctgttg 2040
cggcaacgcc aaaaatccac cccggtggcca acaaaatctct ctgctcccgt tttttccacc 2100
ccctgctgca cggcagcagcc cttggaactg attaactttc cctccgagct gtggcgatt 2160
aaaaatatcg agataaacog tggcctcaaat ggggttataa cccgcaacca gatgggcatt 2220
aaaaagagag ccggctcacaag ggggtcattt tggcctcaca gcccatacttc ttttctccgg 2280
geccattcag cgggaaacaattgt gcgtctcccc cttgctcggc aacctgtgcttt 2340
ccctgctgcc cggactccac cttcggggata ctttttatttt cctccttttct 2400
aaacggtgac ccggcactgt aaaaaagcctg ttagtaaatc ggtctctaaa aagccgaa 2460
aggtcacact gattattttg gcttctcgac aacgttcgcc cctgctccgt gtttttccctt 2520
aaggtatgcg acggagtgca tctatttcct ttcctccgact gtatctttcct 2580
ctggtttttg cgtcgaagca gaatcacc ggtatcgcg ggggatccct agagtgcacc 2640
tgctgctgcag cctgttttta cccgtggtgc cggctgtgcgt cagatgacat gatattgcct 2700
aatcgtgcat accctatttg tagatgtttt aactggcttt ggggccattt cccctccctc 2760
cctcttcgtg ggcgatcata ttaagtcgaa atctgttttt cttgtcgttac 2820
tatggcctac aaaaactgac atagctacac aaatcttcta aatgctact tttttttcct 2880
gccttctgtc tggggttttt tccacactct ctagctactt ttcctgctatg gcgtctaatg 2940
ttaataattg cgggttataa ggttctac aagcttcttttt ttaacacttga aagatgctag 3000
agggcaggg ggcgcaaatc ccttataaat cccagaataa ggcgcgattt ggggtggttt 3060
ctgttcttag cttgggaacag aggctcagct taaacacgct ggacctccaa acctgaaagc 3120
gaaaaacagcg caacatggag ggtgctcgca actatcgcagc aacgtgtaataaaaaa 3180
tggggtgtaga gttggctaaaa gcaactaatc ggcacaccg gggagagccc cgattagct 3240
cctcgggggg aacggcggccaac ggggagagctt gggagacagag gggagagccc 3300
ggccttggtg acggctctcg ggtggtgattt cagccttgtcc cggccagccc 3360
aatagctacg cggcagcagtcc tgcagcctccta gtcgtagctc aacgtacaca cgcgcgtcgc 3420
catattttttt atttttttta aatcattcaca atatgtgcac gttcactcagcta caataacatc 3480
gatgtagcgg caatataatt tgaaaaaagga gatgtagttg ggggaaagct aacagctgtgg 3540
gaatgttctca tcaattgcgt tggggagaaaaa acaagagccag ggttcttgtag 3600
aagcgtgcat etcaattgag cagcaaccag gttgggaaag tcgccaggct cccagcagg 3660
cgagaatagc caaaggatc atctcaatta gtcagaacc atagtccccg coctaaacct 3720
gccatcggg ccccttactc gcccccagtc ccgccacttc ccggccccag gcgtgactat 3780
cccccttatt tctcagggc cggagccgccc cgggtctcct gcgtttaccc cagataggctg 3840
agggagcttt tttgagaggcc taggccttttg caaagatcga tcaagagaca ggtgagggag 3900
cgttctgcct catctgaaacc gatgaggtgac acgcagttgc tcggcccgcct tgggtaggaa 3960
ggcattcgag ctatgactcg gcacacacca caatctgctg cttcgaagcc gcgtctgttc 4020
ggggtcagc gcaggggccc cccgtttcttt ttctcaagac gcacctgctc gcgtgctctga 4080
ataaacgtcaca agcagggcag gcgggctctat cgtggtctgc ccaacgggct gttctctgctg 4140
cagctgtctgt cgcctgtgct gctgaagccg gaagggccttg gctgtctgatt ggcgaagttgc 4200
cggccgagag aacgctctgc tcataactcg ctctcgccga gaaagtactc atcattgtgc 4260
atctcatcgcc cgggactgtc acgtctgtac cgcctactcg cccatctgcg ccacagcaca 4320
aacctgctc agcgcgagcag cgtactctga tggagaagcg tggggtgagc caggagatgc 4380
tgaggcaaga gcatacaggag gccgggcgct cccagcagtc gcggccgctc aagcgcgacg 4440
tggccactcg gcagggcgccg gcggctctgc ggtgcgggtc gcggcggcgc gggggaggtc 4500
tggaaactgc gcgggtctct gcgtttactc gcagctggtc gcggggtgtg ggagggctgt 4560
atgagcacgc aggggtcggg ctgcctgtga ttgctgagaa gggctggtga gttgggctgg 4620
accgcttctc gcgtgcctac gcggcgcgcc ccccgctagt ggcgtttctc 4680
gctttgtagc cgggtctttc tggagcggagc ttcgggggct taaatgtagc aacaagcag 4740
gcccaacctg ccatacagag aatcctgatc cccgcgcgcc ttctatgaaaa gttgggcttt 4800
cgggtctgt cttccgaggg gccggtcttggt gctctccagc gcggggtgtc ccagggcgtg 4860
gccctgtgc gcacccaggg ggggctcaac ttcacggagc aagggagcac tacgcgaag 4920
aacccgctgt atgcagcgcac taaatagcgc gattactggt gttgggtctt gcgggctgt 4980
tctactagcg tggggctgct cccagcggcc gcgttcgtct gcacgcactc ccacgccca 5040
tgggcccaac tacgcggcgc cccctccttc ccccccctcgc ccccgcctgc gttcgggtag 5100
agggcagcgcc ctcgcagcg ctcggcgggc gcagggcctct gcatactgcgt caggttaact 5160
atatatcctc taggtgattc taaaacctca tttttattt aaaaagatct caggtagatg 5220
cctttttcag aatctcatag ccaaatctcc ttaacgtagc ttcttcgcct gcacggctgtc 5280
agacccgtac caaagactac cggctctgctc tgtgggtctc gctataactgtc 5340
cgttgtccaa acacaaaaac cacggctcag gcgccgagcgt tgggtcaagcct tacaagact 5400
aacgtctcct ctttgcaggg taattggcct cagcagacgag cagatacctca taatctgtct 5460
tctagtgtg cgtagttttag gcaccacact ctcagacact ttgacagcgc ctactatact 5520
cgctgctaca actctgttgc cagttcgtgc ctcagcgtgc gattaggtgt gttctacgag 5580
gctgggtacg cagacagctg taccggatag gcgcagcggc tggggtgtaaa cgggggcttc 5640
gtgcacacag cccagctgttc gcagggcagc tcacacgccg ctcgatatac tacgcggtgc 5700
gctgagacgc gcgggtcggg gcagagcggcg gacagtttgct cgcctagcgg 5760
caggggtagc acagcgagac gcagcgagga gtctccgagg ggaacgccct ggtgctttaa 5820
tagtttcttc gcgggtcggc atccctgcct tgcacgctaga tttttgtgat gttgctgcag 5880
caagttcgcag cgtcgccaga ggcgtccaca caacctcactt gccttcagtc ggctggcttg 300
tgtgtgtatg ggcgtcgctg gttgctccct gggtgcgcgg cattcttctc tggaaattgt 360
gacctttttgc aacctctctgt ggagttttttg gacctcacttt gatgtgcctgt ggtgcaaggg 420
cagactctgg acctgtgcgg ttcattctgt tcggtctacag tcggcattta ctctctcttc 490
caatgcacgg agcgctgctg ccaagaataa ggcccggttg atcatctctga tgctgtggtgat 540
tgtgctcggc ctctgctttct ttggtccact tctgatgcac tggtcaggc ggcggcagca 600
ggaagccacat aacctctgag ccaagagcc ctcggctgtcag ttccttcctag accggcctca 660
tgcaattgcct ttctcctcgc ctgtctctca cgttcctcctg tgatatcgctt tctctcgctt 720
tcctcggcgt ttgggaggc cccaaaaagg gctccagagc atggacaaat ctagggcccgg 780
tcctccagtc ccaagatcctt ggcgtgtaag ggagagggcct ggaagttttg atggacacgg 840
cagacccctc agatatcgtg tgaaggggca ccaagcccttc gaaagcttag gcaatcctat 900
ggggctacctgc accctctgtc ggtcgccctct ccctcgtttt acatctctgc atgtgatctca 960
ggataacccc atcgtcagag cagtttacat ctctttcact tggatagggc atgtccacttc 1020
tggttcaaat cccccctctct acgcggcccc cccaggtgctt ccgagttcgt cctccagggct 1080	tgtgctccgt ggcaggtttc ctgggacaat ggcgtaccca gcaagggcag 1140
cacaggaggg cacggatgct atccaggtgga acggagaaag gaataaaagc tgcctgtgtga 1200
agacocccac gcggccagag agctttgctgg ccaatcaaggt acgtgtcact ggaataacatt 1260
tgatccagc gggagcaggt tgaagtacaa gtttcctcctg ctataataag gatccaggccgg 1320

<210> SEQ ID NO 16
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<222> OTHER INFORMATION: Synthetic oligonucleotide sequence

<400> SEQUENCE: 16
aattgcattc tcacgatgta tggatggtg agttgcagag gaccccatct tg 52

<210> SEQ ID NO 17
<211> LENGTH: 1320
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 17
tgcaaccttc ctaatttttt tcataccggc gtttgggggg ctagcaggag gaattcacc 60
tgagttggtg gttggcccaa gggaatata tcacccctct aaataactcg attgctgtta 120
cacgggtcac caagagacac tacacgctca atagctgtcgc agggcggcgg caggtacagg 180
atgcagaggg ggtgtgagac ggtctcctca cggagcagag aacagcactcc 240
tcagctctgc caagggcagc aaggataagg gtcgaggtga gatctctctct tgacagcgtg 300
acggcagcgg gtcggtgattc tcaagggagc gctgggacact gctgcagcagc 360
ttttcgctg cttcattgct ctgcgctcgc tcaggggaac gggcagcttc 420
agacccagaa caccgctggt acactgcatg caggttctct tctaaagaga aacgagtgtg 480
tctctgtag taactgtacaaa aagccgttgg agtgcagcag gttggccota aaccaagattg 540
agaataattt gggcagcagag gactgcagga ccaagactgt gtggccctctg tggatttcctt 600
<210> SEQ ID NO 18
<211> LENGTH: 1380
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 18

ctatacctgg tttttgggcc tgaacgaggg aaatcccgt gcaggtgac atgggaacttg
60
tgcctcacct aagggacaggg gagaagagat atgtgtgcgt gctccacagga aatatccg
120
acccctaaa taattcgatt tgcgcccaaac gttgccctcc ggtgacatg ctttacatg
180
aagtgcgaag gccaagggag ggaaaacttg cttgagggcg agatcaaggg ctgctcaccg
240
ttccctgaga cctcccttcc caaggggcat gcgtgacccc atgggacagc caggtgggcc
300
aggtggagat ctcttcgccc cacccggcgg cccggcggtgtc cggcgcgggt gggggactg
360
agtacccgca taataggtg gaacttttta tcaagtgtcct caaattggac cctgtggactga
420
atggagccgg gccctctcgg ccagtccagg caaccaagac ctagtgccac ccgctgtcag
480
gttctctctag aagagagaca gagtgtgct gttgagtttt tctgagggta aagggggagt
540
gcaaacagtt gtcgctaccc caatattggg atgttaagcc caaggtggag ttcgggaggagt
600
cagtggtggt gctccgctggt attttttgc ggctttgcct tttatccctc cttctcattg
660
gttaaatgt tggattcacc caaggtgag ccaagtctta ctctattgtg tgtgggaat
720
cgacaacgta aaaaagaggg gagttggaag gacactac acgaagccct gcccaaaaaa
780
cagttgctgc tccactcctt ggtttgaccc ctacccctgg ctggtgacct gcgttcgctt
840
caccctcag cctccgtgac acctattccc cggattggct cggcctccct gcaccagggg
900
gcagagaggt gcagccacag tatacgagcc cttgacccct cctttgacca gcccctcgtg
960
cgccaccc ccccaacctc cttcaagaag ggagagggc ccgcccaaca ccccaagagc
1020
tggagactg tgcggccggt cctgcctgg cagctgcttg gagctgcgtgt cggcttggct
1080
gacaggggag gctgtggcag gacatctgct tggagcttg ggaagctggg gacggtggct
1140
agagggggcc tgcgggagcc gcacgaggct taggagctgt ggaggcagct gggcccagca
1200
gccgcggcg gaggccagcc ttcggggtct cccgccccgc ccgccgcccc ggccggcccc
1260
tgagagggagt gacagacatt gcggggcttg tgcagggggt gagaagttgc gcgggcgggtg
1320
-continued

ccaagtcctt cagataatga ggtacgacg gagaatccgt aatcattgtg ctagactgtt
<210> SEQ ID NO 19
<211> LENGTH: 780
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<400> SEQUENCE: 19

gttttcttgg gctgcagag gaatccatg agaactgaa gctgatccg ggaagtggag 60
cagcccaag agacgtcacc caagagaca ggggggcccc aggcttcacg gcgttcttg 120
ttcctcagc tcctctctct cgtgactgtg gcaagggcga ccaagctttt ctgctctgtg 180
cagtttggag tgtgagcggc ccagagggaa ggttctcctc tctaatcagc 240
cctctggccc aggcagctac atcattcttc cgaaccgccg gtgacaggg tgtagcccat 300
gttgtacca acccccttgc tggaggccag cttcaagttg gtaaccgccg ggcacatgcc 360
cctctggccaa atgaggttga gctggagatg aaccagcttg tgggtcctac agagggcctg 420
taactaatt aatcctaggt cctctctgag gcggctaagg gcctcccctt ccaatggtctc 480
tctcaacaca ccattctgctc catcggcttc tctctgttcc ccaagactca ccctcctctc 540
gccataaaga ccgggagacc ccagaggggg ctggagcaca gccctgttat 600
gagccctct atctgagagt gtttctcctc gttgagaggg tgtgacgact cagcgctgag 660
atcagagcc gcagtattct gaaatctggc gatgctcgag tggagtata tggagctatt 720
gcctgtgtat agtcctggcc cgcgggcccc ggtacctccg gatctgata actagcatc 780

<210> SEQ ID NO 20
<211> LENGTH: 300
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 20

gttttttttg gctgcagag gaatccaccc atgttaccat gaaacttgggg aatgaactgt 60
tatctctgt agcgggttta ccctcccctcg cegttattct gttcaatatg gactctgaat 120
gtatcctgct ccaagagagg tacgctcctc atcggtttgt tgtgcttat attgaagcat 180
tggacaaga tctagctgacat tgtgtggttg gttctacttg gcgggagatg cagctacgcag 240
acgcagttgg gttggagatct cgtcataatg cttggcctgg gcgggcccc gttacccgat 300

<210> SEQ ID NO 21
<211> LENGTH: 1620
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 21

gttttttttg gctgcagag gaatccaccc atgaaacttg ggacagcact gtttaatttg 60
atagctgctct tctcctcctg cgccttttca atgttctcct tctacatggt ggtttctg 120
cagcagctct gcagttgatc accgctgccc gctactgggt atcggccacc ggtgaatggg 180
aaaaatctgga tctacggtta gttacgcttg tgtttttctt gggggaaaa cttggcctgg 240
cagcaggtgg tctacacacct caataatggt gcgccacgagataaattcag gcgggagcac 300
-continued

gatgcaaatg ggcgtgcgac cattgcaatta accaaacga cagatggcgt gcaggtgtagc 360
acagcagaa tggagggcat acggcagagt gttgataaccc actttgaaat ggtactatcc 420
gggggaatc atccactacct gcacaggtccta cccactcctta tcgttcgtaa tggctctaatg 480
gttcaacca tcaagtgtaga gttgaaagat acacatgaggg accgcagccgg tgggctcgaat 540
gttgcttttg acacaaacct agggcaattt ggctgattca cggatcaca atggggccact 600
tatagcgac catcagccag taactaagt cgggtagccaa catttaaggtt gaaattggtag 660
ggggctgtgct ccaagtggtcc gcaggtgaag gttatatttca cgggcatctcc tatccagat 720
gtcgtgctgct ccagtttccac cgtctccaca cccgatatct tcgtgtgtag gcagatgtcag 780
tccacactta cttcttctcc tcctgataag aagggccatt tttaagttg tagtacaagggc 840
ttgagatgtt cttcacaagct gttgcaaggt agtatattg ccattatacgc gcaagccagat 900
agctatacgc gcaaggtgtgt tggaatagtag tgcggtgatg tcacaatcgc gcgcaggtt 960
gatcctcctga tactagatgc attgcaaaaa aaaaatccccc tattccccgggc acctacgctg 1020
acggtaatcg tggtaaagc gcacaaatctc atctgggata aagggctccc gaaaaagatc 1080
ttttaaagc cccacatcca gttcagatt gataaagatt gtctaatatg aactccagat 1140
gactggtgtct gtctactac ttccgcctattc cttacctagc attggaagtt gttgacagt 1200
aacattttc ccctagagga atggggtgtag agctggaagga gtaaaaaatt cccagtttat 1260
tcgggtgatt agtagtttca cccaaatcgg tggatataag atggggccag atggtctgtaa 1320
ttcagtcttg cgcacagcag acatgtccca ggttgcatgc tgtctgctt cttcagttcat 1380
tcgctgctaa cccaaagcag gcgttgccct gacggtgagat tgggggaagta gttggggagt 1440
tgcaaggcct atagttctga tggccatctt gttgaaattt ggtgctaaaaa gaccagcagc 1500
gattttgac cctgtgatag gcgacacgcc gcagctgacac cagggtgcctc atctcttgccg 1560
ttcgcgcctct tggtgctgtc aataatacg tcggctacgc aagttggtccc ggggccccgg 1620

<210> SEQ ID NO 22
<211> LENGTH: 208
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 22

gaattcagc gtttttttaga cttggctgtaa tgaatcaccac caggatcaca ttctgcagat 60
gtcgaagcgc gcagcattag tcgccttggt ttatatatttt aagttatattg tacatgagg 120
aaataaagtg aacaaacagc ctattgcagt gcacatctca cccgtaactctg ttcccccttg 180
gacacaagcc cggacacaccc aatcttga 208

<210> SEQ ID NO 23
<211> LENGTH: 1546
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 23

gaattcagc gtttttttaga cttggctgtaa tgaatcaccac caggatcaca ttctgcagat 60
gtcgaagcgc gcagcattag tcgccttggt ttatatatttt aagttatattg tacatgagg 120
aaataaagtg aacaaacagc ctattgcagt gcacatctca cccgtaactctg ttcccccttg 180
gacacaagcc cggacacaccc aatcttga 208

-continued

tactgcatcc ggggtgcctc ggcgtttaaa cgggtgatcag actgcggctc tgcgtgttcc 300
tttagagt aaacctgca aaaaatatt tgtgtgttatt ggcgtgaggg tgggggactc 360
ggaattact ggcgcacgta attatcgag aaggtcggggc ggttttttaa aaggtatataga 420
tgctttacgc ctgatgcggc ctactacacgct cgtgtctctg ctaaaaaaaaaa cgggaacacc 490
ggactacaag ccaccgctgg ctcgcatcag caccggtctgg tcacacggtg tcacaccactca 540
tacgccgccc tgcggcgtcg acctacaag aaaaatcaca ccaacggattc tgggaatggc 600
aacaacgca ggtctgcggca cccgtaacgt ttctaccgca gagttgacag atgcacgcc 660
cgcgggggct ctggacaccct tgtgaattcg ccaatgctac ggtgcgagcc cgcacagttg 720
aatactcgcg ggttaacgct tgtgaaaaagg cgggaagaag ctgattaccc aaccggtctg 780
tacgcgtgt cgcgcagttg cgcgttcggc cggcgcaaaaa acctttggtct aacgggacc 840
cgtcggtgaa tgtccaggaa aaaaagtcgct tgaacaggca cagggcgcgtg gttacagtt 900
gggtgcagct gttcgccttac gtaaaccgct ggcgcagacg atccacgaaaa aacccctctg 960
tgggtcgttt ctggtacgcgg atagatcccg gcgggtcgtct ggcgcggaag caaacggtca 1020
tggcatactc gataacggc cagtcaactg tccggaacct cccgcaacctg atgcacagtgt 1080
aaccacccct ggcgcagatg cccgacaaac caccgaattg tgtgaattag aatgagaagg 1140
ccttttcctc ctaagttggaag gtgtggttaat ctgataaacg gtcacatgct cgtacccctg 1200
tgggcaccat gcggacagcc tcgagatcga tgtgacgctc caaaccgcgc gcgaaccaggc 1260
taaaagagc ggtcactcag tgggtcatag cccgcgggtc ccagccatc cccagcagat 1320
ttgtgagcgc gataacgcct ctcggcgcct caccgcgcgg ctaataccaa aagatgcggc 1390
agttgttgga atcgctagct ggaatcggag aaggtttttga cgaacactcata cgggctgca 1440
gttgcttgt ggggttcgct gcggctagtc gcgggagct ggaggactg cgggaggcgc 1500
cgatccttt tcacactatg aacgccgtct ggggtgaaa tccaga 1546

&lt;210&gt; SEQ ID NO 24
&lt;211&gt; LENGTH: 148
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: E. coli

&lt;400&gt; SEQUENCE: 24
gaaattcagc gttttttataa cttcagttaa tgaattcag caggtacaca ctgctctagat 60
gaaaatataa aacagggcgac gcucctgctc aattacgca ctaaaaggca tgaatttttc 120
cgcgtgcgcc ctgcctttaaa tcccttctag

&lt;210&gt; SEQ ID NO 25
&lt;211&gt; LENGTH: 1174
&lt;212&gt; TYPE: DNA
&lt;213&gt; ORGANISM: E. coli

&lt;400&gt; SEQUENCE: 25
gaaattcagc gttttttataa cttcagttaa tgaattcag caggtacaca ctgctctagat 60
gaaaatataa aacagggcgac gcucctgctc aattacgca ctaaaaggca tgaatttttc 120
cgcgtgcgcc ctgcctttaaa tcccttctag
<210> SEQ ID NO: 26
<211> LENGTH: 3840
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 26

```
accatagccttgtaaata ccgcaacagt gcagcaagag aasatacgcg atcagggccg
atcagcgcaca gcgcgcaca aaccttctcg actactaataa gttatgccg tataattgcg
ccggctctt attggctttc ggatggcttg tgggatgtgt tattatagtt gattatagtt
gatggttggc accacgccg gccggttcg gggggggggt tggggggttc gttggggggg
```
-continued

atgagacgct cgctgaaacg acgcattcggg aacgccggtta agtcatcaggt catcgacaggt 1140
ctgtgtaag ctaacagcat gacgctgagc gctgctgctg gctgctgctg tcgctgctgctg aatgacgctg 1200
cgctgctgcc gcgtggcagc tctgcctgcc ggcacggaga aaccagcgctct ggaactgccg 1260
cgcagacttg tcagcgttcag cgtgatacgtt tttggtctgct gcagccgacaa 1320
gttacgtgaaa cccagagcagc gaaagcggac gaaagcggac gaaagcggac gaaagcggac 1380
atgctgtaag cctgagagag cctgagagag cctgagagag cctgagagag cctgagagag 1440
cctgccctgg acacgccaga cgcctggcag cgcctggcag cgcctggcag cgcctggcag 1500
aaacgcagcag cgaaccagcag cgaaccagcag cgaaccagcag cgaaccagcag cgaaccagcag 1560
ttaaatcctc gacgctgctg cctgccctgg acacgccgag naacccgag cnaaccgag 1620
aatagtcgagc gctgtcacagc gactacggag gaaacccggag gaaacccggag gaaacccggag 1680
ccccccctgg aagcccggag gctgctgctg gctgctgctg gctgctgctg gctgctgctg 1740
gtaagacgct cgctgaaacg ctcgcctggc agccttcctg aacccagcagc cgcctggcag 1800
taaatcctc gacgctgctg cctgccctgg acacgccgag naacccgag cnaaccgag 1860
tggccctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag cgcctggcag 1920
tgtgctggctc gctgtcagagc gactacggag gaaacccggag gaaacccggag gaaacccggag 1980
cgcactagcag tattaagctgct acgcctggcag cgcctggcag cgcctggcag cgcctggcag 2040
agcgctgctg ctaagacgct ctcgcctggc agccttcctg aacccagcagc cgcctggcag 2100
cactgtaagc cgtgctgctg ctaagacgct ctcgcctggc agccttcctg aacccagcagc 2160
aagaactgttg ctgctggcag cgcctggcag cgcctggcag cgcctggcag cgcctggcag 2220
ccccctggc acacgccgag ctaagacgct ctcgcctggc agccttcctg aacccagcagc 2280
cgcactagcag tattaagctgct acgcctggcag cgcctggcag cgcctggcag cgcctggcag 2340
gtggccctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag cgcctggcag 2400
cgcctgtaag cctgagagag cctgagagag cctgagagag cctgagagag cctgagagag 2460
gcgagctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2520
aagctgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2580
tgcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2640
gctgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2700
tgcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2760
gtcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2820
tgcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2880
tgcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 2940
tgcgctgctg ctcgcctggc agccttcctg aacccagcagc cgcctggcag cgcctggcag 3000
agggagatcag ccacccgctgttggctctgtaa cataaaacaaacctt ggggtgagcag 3060
aataagccag ctaaaaagctc gtaaaagttctg gtaaaaagttctg gtaaaaagttctg gtaaaaagttctg 3120
ccagcctagg ctggtctgctac ccagcctagg ccagcctagg ccagcctagg ccagcctagg 3180
ctgccgctctg aataagccag ctaaaaagctc gtaaaagttctg gtaaaaagttctg gtaaaaagttctg 3240
ctgccgctctg aataagccag ctaaaaagctc gtaaaagttctg gtaaaaagttctg gtaaaaagttctg 3300
aataagccag ctaaaaagctc gtaaaagttctg gtaaaaagttctg gtaaaaagttctg gtaaaaagttctg 3360
-continued

gcttggtgc tggtggtggt gtgcgcgctga tccgcgtcag gcctaaacct gcgtacgcgtc 3420
gttgctcagaa cgaagaaccc aacgtgggta tcaagaggt gcctgctgca atgggaagtc 3480
cgtgcgctca gatcgatatt aacgtgcgag aagaacctga tcggtggtct aacaccctta 3540
aaggcgccga cgcacaacgc ggctacaacc cagcacaagg aacatacgac aacatgtgac 3600
acataaggat cctggtacca aacaaagtaa ctcttgcttc ctccgcatgc gcgccttcgt 3660
tgcctgtgct gcctgctaccc accgaatgca tgcctagcag cctgccgcaa aacgtagcag 3720
cctacgctagg cgtgcgcgct gcggaggtgag gcctggtgag cagtggcgcg atagtgtaat 3780
ataagccgct gatcgcgtcgt gtcgcgctct agaggtaccc cgggtacgc gcgcgactcc 3840

<210> SEQ ID NO 27
<211> LENGTH: 480
<212> TYPE: DNA
<213> ORGANISM: E. coli

<400> SEQUENCE: 27
gaatcaggg gccttttata gcgggtctgaa tgaaatctcg cagggatcaca ttcctgcaaat 60
gatcagggc gccttgctac gcggagacga taaatattt ccctgctgct gcagacagttt 120
tgacaggt gcctctcagc cggcgcggcag gatctgcgtg gattgtctggg caagtggtcg 180
cgtgcgctga aaaaatatcg cccggattct gcgggaatc gttgaagct gcgttggcagg 240
acctgaacgt aggacgtgta aacactgatt aaacgcgatt gcctggtcag gatggaaggct 300
cgtgcgctga ccctcgtcta gcgggtctga aacaggtgaa cgggtacgcc ccaagtgggg 360
tgcctggctt aaggtgcgct tgaagagatt tcggctgagct aacgtgccccgt gcacagtata 420
aatagctgcat ctctgctcaag cgggttagaa aataacgtaa tcgatgctta tgaagagatt 480

<210> SEQ ID NO 28
<211> LENGTH: 1140
<212> TYPE: DNA
<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 28
atggcaacca cgcagcggca ggggccaccc cggcttctgg gggaatatctg ggtccgggaa 60
cattatgatt aaggggagag gcgtggcggcg aggctgcctga ggtgccaccc 120
ctccatcagc ccctctccct ttcggtcacc tctagcgtca tcggcttggga ggttcgtgct 180
ttcgggatt ccctctggaa aacaaataaa tttcataacc gcgttcttct ttttcctggc 240
aaacctggtgc gttctgacgc gcgggtccgg cggctgcatt aacgctcaca ccggttacccctct 300
ggtttagga cggggtctga ctctggcacc gcggggttct gggtgggagg cgggttctg 360
gttctggtg gctgcgacac atgcgctcta ttcgctgcgc gcgggttttg 420
atgggtgaag ggcgtctgct gcgggtctcag gcgggtctgct gcgggtctgct gcgggttctg 480
atgggtggtc tttgtgcggtc tttgcgtgctgc ggtgtgcgc gcgggtttgggc ggttgcggtc 540
ggaacatctgc cggctcgctc cagactgctc cggctcgttc ccggctcttc cgggtcttctg 600
cgttcagcgc ccctcgtgctt ttcggtctgc cggctgctgc ccggttcgctgc ccgggttctg 660
ttcggtctgc ggttcaggct gcgcgctggt gcgggtcgcgc gcgggtctgct gcgggttctg 720
cgttcagctg ccggtctgct gcgggttctg gcgtgcgctg ctctgcgtgt gcgggttctg 780
ttctgtctgc gcgtgcgctg gcgggttctg gcgtgcgctg gcgggttctg 840
cagtggttca tacagcgc gcgtactgct cagctgcgtc acctaggcgc ctacgctgct
900
gcagcaagg atagcagcgg ctgctctctgc cggtgtggtg ggcgtgtgct ggtgaaaggcc
960
aaggggaccc aagcgcctccgc gatgagccc gctgctgacc gaggcagaaatgaaatcagc
1020
tccagtacaca cagcagcgcc caaacttcca aaggtcagag aagacctgcc ccagtgtgct
1090
acctctctct gcgtcactgga caaaaagagctgctccaga atggggtctc ctgcaagtga
1140

<210> SEQ ID NO 29
<211> LENGTH: 379
<212> TYPE: PRT
<213> ORGANISM: Rattus norvegicus

<400> SEQUENCE: 29

Met Ala Thr Thr His Ala Gln Gly His Pro Pro Val Leu Gly Asn Asp
1  5  10  15
Thr Leu Arg Glu His Tyr Asp Tyr Val Gly Lys Leu Ala Gly Arg Leu
20 25 30
Arg Asp Pro Pro Glu Gly Ser Thr Leu Ile Thr Thr Ile Leu Phe Leu
35 40 45
Val Thr Cys Ser Phe Ile Val Leu Glu Asn Leu Met Val Leu Ile Ala
50 55 60
Ile Trp Lys Asn Asn Lys Pro His Asn Arg Met Tyr Phe Phe Ile Gly
65 70 75 80
Asn Leu Ala Leu Cys Asp Leu Ala Gly Ile Ala Tyr Lys Val Asn
85 90 95
Ile Leu Met Ser Gly Arg Lys Thr Phe Ser Leu Ser Pro Thr Val Trp
100 105 110
Phe Leu Arg Glu Gly Ser Met Phe Val Ala Leu Gly Ala Ser Thr Cys
115 120 125
Ser Leu Leu Ala Ile Ala Ile Glu Arg His Leu Thr Met Ile Lys Met
130 135 140
Arg Pro Tyr Asp Ala Asn Lys His Arg Val Phe Leu Leu Ile Gly
145 150 155 160
Met Cys Trp Leu Ile Ala Phe Ser Leu Gly Ala Pro Ile Leu Gly
165 170 175
Trp Asn Cys Leu Glu Asn Phe Pro Asp Cys Ser Thr Ile Leu Pro Leu
180 185 190
Tyr Ser Lys Tyr Ile Ala Phe Leu Ile Ser Ile Phe Thr Ala Ile
195 200 205
Leu Val Thr Ile Val Ile Leu Tyr Ala Arg Ile Tyr Phe Leu Val Lys
210 215 220
Ser Ser Ser Arg Arg Val Ala Asn His Asn Ser Glu Arg Ser Met Ala
225 230 235 240
Leu Leu Arg Thr Val Val Val Ser Val Phe Ile Ala Cys Trp
245 250 255
Ser Pro Leu Phe Ile Leu Phe Leu Ile Asp Val Ala Cys Arg Ala Lys
260 265 270
Glu Cys Ser Ile Leu Phe Lys Ser Gln Phe Ile Met Leu Ala Val
275 280 285
Leu Asn Ser Ala Met Asn Pro Val Ile Tyr Thr Leu Ala Ser Lys Glu
290 295 300
Met Arg Arg Ala Phe Phe Arg Leu Val Cys Gly Cys Leu Val Lys Gly
Lys Gly Thr Gln Ala Ser Pro Met Gln Pro Ala Leu Asp Pro Ser Arg 325 330 335
Ser Lys Ser Ser Ser Ser Ser Asn Ser Ser Ser Ser His Ser Pro Lys Val 340 345 350
Lys Glu Asp Leu Pro His Val Ala Thr Ser Ser Cys Val Thr Asp Lys 355 360 365
Thr Arg Ser Leu Gln Asn Gly Val Leu Cys Lys 370 375

<210> SEQ ID NO 30
<400> SEQUENCE: 30
000

<210> SEQ ID NO 31
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 31
aattgcagctccacagctccagttgtta 32

<210> SEQ ID NO 32
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 32
aattgatctctatgctacagtctttta 32

<210> SEQ ID NO 33
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 33
aattgatcccagaagctttttctctcttttta 32

<210> SEQ ID NO 34
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 34
aattgatccccgaagcgcggctttttgcttttta 52

<210> SEQ ID NO 35
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 36

```
ttatggcaac acgcacgcgc cagg
```

<210> SEQ ID NO 36
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 36

```
agacgcgtac ttgcagagga c
```

<210> SEQ ID NO 37
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 37

```
asattgtac agcgccgga cgggcaacc gc
```

<210> SEQ ID NO 38
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 38

```
asattgtacc tcacttggag aggaccccat tcgt
```

<210> SEQ ID NO 39
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 39

```
asattgtac agcgccgacu cgggcaacc gc
```

<210> SEQ ID NO 40
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 40

```
gtgcgcacc atgttgagca a
```

<210> SEQ ID NO 41
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 41

```
ttaaggatcc ttacttgtac acgtgctca t
```

<210> SEQ ID NO 42
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 42

```
ttaaggatcc ttacttgtac acgtgctca t
```

<210> SEQ ID NO 43
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence

<400> SEQUENCE: 43

```
ttaaggatcc ttacttgtac acgtgctca t
```
<210> SEQ ID NO 42
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 42
ggtcgcacc atggtgagca a 21

<210> SEQ ID NO 43
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid primer sequence
<400> SEQUENCE: 43
ttaaggatcc ttgtcacgc tcgctcatgc c 31

<210> SEQ ID NO 44
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 44
ccaatggac ttaccaatga cgegg 25

<210> SEQ ID NO 45
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 45
gttgctta gcaggaatgc tggg 24

<210> SEQ ID NO 46
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 46
cggcgctga gatgtttgaa ccaatggac ttaccaatga cgegg 45

<210> SEQ ID NO 47
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 47
ggcctctag attattaatc agcttgttta cgcgagaatg tggg 45

<210> SEQ ID NO 48
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 48

gctgactgg gcgttttat ggacagcaag c

<210> SEQ ID NO 49
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 49

gcgttaaatt tcgagaaga ctcgtcaaga aggcg

<210> SEQ ID NO 50
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 50

gcgcctactg acgtagtcg accgctggac tagcgaagtt cctatacttt ctagagaata

ggaaccttcgc tagactgggc gcgtttatatg aggcaagcg

<210> SEQ ID NO 51
<211> LENGTH: 109
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 51

cagagttgtt tcgcttttgtc tgaagttgata cgcttttgtg gaagtctcota tttcttagaa

tagtagagca cttcgcttta ataattcaga agaaactgcg aagaagcg

<210> SEQ ID NO 52
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 52

cggttaccaat tagacaaact tgcaggg

<210> SEQ ID NO 53
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 53

ttaaccttttc tgcgaattga gaagtgcgc

<210> SEQ ID NO 54
<211> LENGTH: 28

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 54

ttagtgtaac acttcaat tgcagggg
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 54

gtgcagctgattgctttgagtgcagccag 28

<210> SEQ ID NO 55
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 55

gctgcattg cgggcttca gttcgcgctg cатаctgтс сгттacсаat тагаакаact 60
tcagccg 66

<210> SEQ ID NO 56
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 56

gctgcattg cgggcttca gttcgcgctg cатаctgтс ttaatctttc tgcgaattga 60
gatacgc 69

<210> SEQ ID NO 57
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 57

gctgcattg cgggcttca gttcgcgctg cатаctgтс ttaataaagt gагtогатat 60
tgtctttgtgt gaccag 76

<210> SEQ ID NO 58
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 58

gctgcattg cgggcttca gttcgcgctg cатаctgтс сгттacсаat тагаакаact 60
tgacgg 66

<210> SEQ ID NO 59
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 59

gctgtttcgt gaaaaaaggg gctgctcaaggg 30
-continued

<210> SEQ ID NO 60
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 60
gcgctttca tggtagagaa gagatccg

<210> SEQ ID NO 61
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 61
cgcgctgca gatgcgtgtcttagaaacc ggcgctgctca ggc

<210> SEQ ID NO 62
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 62
gcgctctag attatttt cagococaga gcgctttca tggtagagaa gagatccg

<210> SEQ ID NO 63
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 63
gtcagctcg agacatatag tagc

<210> SEQ ID NO 64
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 64
gcgctttca tggtagagaa gagatccg

<210> SEQ ID NO 65
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 65
cgcgctgca gatgcctcgg ccgagactta tagtcg

<210> SEQ ID NO 66
<211> LENGTH: 58
-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 66

gcgctctcgg attatatct cagccccgag gcgcttttca tgtgtgagaa gagatcgg 58

<210> SEQ ID NO 67
<211> LENGTH: 109
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 67

cgcgctgca gattgacctg gggacgagc tgtttatatct gatagcggtc ttacctccc 60
tgcgattatct gcgctactcg cctgtctcgg aaaaaagggc tgctcagg 109

<210> SEQ ID NO 68
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 68

gcgctctcgg attatatct cagccccgag gcgcttttca tgtgtgagaa gagatcgg 58

<210> SEQ ID NO 69
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 69

gggaattgag atgacgacac tgtc 24

<210> SEQ ID NO 70
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 70

cctggtgat ttcattacag accag 25

<210> SEQ ID NO 71
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 71

cggcgaagct tataataact tgtctgcaat tgtagtcagc ccaactggc 48

<210> SEQ ID NO 72
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 72

cgcgtaatc gccgctgcag aatgtgatcc tgtctgaatt cattaacgac cag 53

<210> SEQ ID NO: 73
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 74

cgcagcgctg ttccttttct cgg 22

<210> SEQ ID NO: 74
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 75

cctcttaaag atataataac tgg 23

<210> SEQ ID NO: 75
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 76

gccgcaagct tgcagcgct gtctttttgc tgg 33

<210> SEQ ID NO: 76
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 77

ccaatgcct ggtcttgtcag gacctcctc taagataa taactgg 47

<210> SEQ ID NO: 77
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 78

gttcttttag cgggaaaag 19

<210> SEQ ID NO: 78
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 79
<210> SEQ ID NO: 79
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 79

gcgagatccc tcgcttgct

<210> SEQ ID NO: 80
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 80
gccgcaagct tcgtcctttag ccggaacgc

<210> SEQ ID NO: 81
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 81
cagagctcga attgtctgtat tcg

<210> SEQ ID NO: 82
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 82
ggtgaattcc tcctgctagcc cc

<210> SEQ ID NO: 83
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 83
ggcgcaagct tcaagcgctc aattgtctga ttgcg

<210> SEQ ID NO: 84
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 84
cgtcaggggt aatctctcct gctagcgc
<210> SEQ ID NO 85
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 85

gcttaactcg agttaataa caacgcgcctgattgtctgt ctctc

<210> SEQ ID NO 86
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 86

gcttaaccgc gggcaacgctgatgcctgcctgcctc

<210> SEQ ID NO 87
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 87

ggcaacgcac gcgcgcgcc ggcacc

<210> SEQ ID NO 88
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 88

cagtcgtgat ggtgatgatg accgg

<210> SEQ ID NO 89
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 89

cgcgcgcgac atggcaacca cgcacgcgcgacggccacc

<210> SEQ ID NO 90
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 90

ggcgcgcgac ctggttaatg gtgatgtgatgtatgacctg

<210> SEQ ID NO 91
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 91
gggcaaccc gggaagcca gcgcgc

SEQ ID NO 92
LENGTH: 30
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 92
gcaagtgtgc atttgtacta caattcttc

SEQ ID NO 93
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 93
cgcgtgctac atgggcaac cgggaacgg cagcgcc

SEQ ID NO 94
LENGTH: 49
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 94
gcccgcgatgc cttatatag cagtgctca tttgtacta aattctcc

SEQ ID NO 95
LENGTH: 28
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 95
ggacctgtgcct ctcacctagg ggacaggg

SEQ ID NO 96
LENGTH: 27
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 96
cgsgasaagac tggsgcgag ggsgag

SEQ ID NO 97
LENGTH: 42
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 97
cggcggtcga catggsactg gtcctcacc tagggacag gg

<210> SEQ ID NO 98
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 98

gcgcgcgtac cttaaatactg aagaagtgg gcgcgcgcgc gcgg

<210> SEQ ID NO 99
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 99

gatagtgtgt gttccc

<210> SEQ ID NO 100
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 100

cgagagac gtcgagc

<210> SEQ ID NO 101
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 101

gggacccact gatagtgtg tgtccc

<210> SEQ ID NO 102
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 102

gctcatactta gttactagag aagacgggcc gc

<210> SEQ ID NO 103
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 103

gagcactgaa agcagtagcc ggasgcg
<210> SEQ ID NO 104
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 104
cagggcaagt atcccaagt agaactgc

<210> SEQ ID NO 105
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 105
cggcgagaat tcatgcgtgaa acagatcgct gcggccagc

<210> SEQ ID NO 106
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 106
ggcgcaagct tcatcagag gccgaatc caggttgac ccagagatgct gtc

<210> SEQ ID NO 107
<211> LENGTH: 110
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 107
tctgatagc gctttacctc cccctcgagt aattcgtcct aatagtgact ctaaatggcc 60
cctggcaca gatcggtact gcctgtgatgc atgctattgg 110
cattatatc aagtctctgt ctccggagtc ctccggatgc tagccaaaca caagttgca tgcatacttg 60
tcaattatat ccaatacgt gcaacaccaa tcatggagc a 101

<210> SEQ ID NO 108
<211> LENGTH: 101
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 108
agttctctgt actgcacatcg ctccggatgc tagccaaaca caagttgca tgcatacttg 60
tcaattatat ccaatacgt gcaacaccaa tcatggagc a 101

<210> SEQ ID NO 109
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 109
cggcggtac catgaacttg gggaatcgc tttttattct gataagcgct ctacttccc 60
tcg 63
<210> SEQ ID NO 110
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 110
ggcgcaagt ttagcgca gtcccacca cttcaggctc cgtactgac atcgctcccc 60
g

<210> SEQ ID NO 111
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 111
tcattcact tgaagctgac cg 22

<210> SEQ ID NO 112
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 112
ttatagctc aggcacaga gcgg 24

<210> SEQ ID NO 113
<211> LENGTH: 101
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 113
gcaggaacct aacctgaact tggggsatcg actgtttatt ctgatacgcc tcttactccc 60
cctcgcagta ttaactgtctt cattcactt gacgcctcacc g 101

<210> SEQ ID NO 114
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 114
cgctgttaac taagcactg caagtattg gcacgacgac agagcg 47

<210> SEQ ID NO 115
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 115
gtcacggcgc agacttatag tcgc 24
<210> SEQ ID NO 116
<400> SEQUENCE: 116

000

<210> SEQ ID NO 117
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 117

cgggtctgca gatgtcaagg ccgacacca tagtcgc

37

<210> SEQ ID NO 118
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 118

cgggtctag attctgtgc tccggcctttt gtcacagg

38

<210> SEQ ID NO 119
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 119

cagcccagc gggcttcca tgg

23

<210> SEQ ID NO 120
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 120

cgggtctag attcagccc cagacgctt ttcatgg

37

<210> SEQ ID NO 121
<211> LENGTH: 106
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 121

cgggtctgca gatgaaaata aaacaggtg cagccatcct cgcattatcc gcattaacga
cgtatgttt ttcgccctcg gcctctgca aatctcttag aacgagg

60

106

<210> SEQ ID NO 122
<211> LENGTH: 106
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 122

ccggtcttag agatattggc gagaagcggag gcggaaaaac tcatctctgt taatgctgat  60
aatgcgaga tgcggtgcaac tgttttatt ttcactctgc ggcggc  

<210> SEQ ID NO 123
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 123
ggtgcaacgca ttcctgcatt atggcg  

<210> SEQ ID NO 124
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 124
cggcatacca gaaagcggac atgctg  

<210> SEQ ID NO 125
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 125
cgctggctgca gatgaaaaa aaaaacaggtg acacgcacct gcactatcc gc  

<210> SEQ ID NO 126
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 126
cgctggcttag aaagcagcgc atacagaaaa gcggacatct gc  

<210> SEQ ID NO 127
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 127
cgctggcagcg ccacataacc gtctcg  

<210> SEQ ID NO 128
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 128

gctggttcat cagctgttg aagtggtg 27

<210> SEQ ID NO 129
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 129

gcgccccgc cttacgacgc acgcgcccaat aacgtttctcg 41

<210> SEQ ID NO 130
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 130

ggcgctcttag attattatatt cgcctgcttg ttgacgct cgtgaaagt gg 52

<210> SEQ ID NO 131
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 131

ggtgacaaac tcgattgcct ttgacggt 29

<210> SEQ ID NO 132
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 132

gccgccactgc ccaccgcct gccccc 25

<210> SEQ ID NO 133
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 133

gctgctactg acgcacaaat cagattgcct ttgacggt 38

<210> SEQ ID NO 134
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 134

ggcgcaagt tattattac ttcgctgcc ccaccgcaccc atgcgccccc 49
<210> SEQ ID NO 135
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 135
ggataaaat tattacctg actgacg  

<210> SEQ ID NO 136
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 136
gcgctgagga actcttctca ctgacc  

<210> SEQ ID NO 137
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220>FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 137
cgcgcgcgctca gatgatcgaa gcccgctcta gactgacag cgataaaatt attcacatga ctgacag  

cgcgcgggatc ttatattc atcatgatct ttataatgc catcatgatc tttaaaatcc  
tcggagcgca ggttagcgtc gaggaactct ttcaactgac c  

<210> SEQ ID NO 138
<400> SEQUENCE: 139

<210> SEQ ID NO 140
<400> SEQUENCE: 140

<210> SEQ ID NO 141
<400> SEQUENCE: 141

<210> SEQ ID NO 142
<400> SEQUENCE: 142

000

<210> SEQ ID NO 143
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 143
tatgttaagg ggtgtgtcac cgccctcagt tagaggtacc cgcctctcag cgaaagggcg  60
tattg                                   65

<210> SEQ ID NO 144
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 144
gatccaatag gcccctttag aagggggtag cgcctcttag actgagccgg tcgacagcc  60
ccttaca                               67

<210> SEQ ID NO 145
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 145
tatgttaagg ggtgtgtcag cgccctcagt tagaggtacc cgcctctcag cgaaagggcg  60
tattg                                   65

<210> SEQ ID NO 146
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 146
gatccaatag gcccctttag aagggggtag cgcctcttag actgagccgg tgcagaacc  60
ccttaca                               67

<210> SEQ ID NO 147
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 147
gatccctaag aaggtgtgca cgccctcagt tagaggtac cgcctctact cgaaagggc  60
gtattc                                   66

<210> SEQ ID NO 148
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 148

tgagaatac gccttctcgag atggagccgg gcctctctag actgagcgcgg tagacaact 60
ccttag

<210> SEQ ID NO 149
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 149

gaccttcaag aagttctgca gcggctctggt ctgaaggtac cgcctctcat cggaaagggc 60
gtattc

<210> SEQ ID NO 150
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 150

tgagaatac gccttctcgag atggagccgg gcctctctag actgagcgcgg tagacaact 60
ccttag

<210> SEQ ID NO 151
<211> LENGTH: 4740
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 151

tgcgcgcttt cgggtgatgc ggtgaaacc tcggacactc gcagctcccg gagaaggtca 60
cagtctgtct tgaagccgct gcgcgggac gcacacgcg tcagcgccgtgg 120
tgctggcttg cgctggtctc cttaactatg ccggcatctga gcagttgga ctgagctgtc 180
acacatgct gcggatcata gcggccagat gcgtgagag aaaaactgcgg atcagccccgc 240
attggcactt cgccgctgcg aactttggt ggcctccggag ggatggtgcct cagccggtt 300
tccgcagct gcggcagaggg ggtgtggtcc cagcgcggatt aagttggtgta aacctgaggct 360
tcccaggtc aacagttgtg aaaaagacgg cagtgccaa gettaattaa tctttctgog 420
aattggaattg aacgaacattg ctgggctgct tcccggttct cccggttaac accaacccaa 480
aattgtgtc atctttcaag ccaactttgg cgtaaatac gacgcagat cggctgctt 540
ggttggagat atattggcag tcggacactc tggcctgtcg cagatattga ttgatggtcct 600
ttccgctctc cagctggaaat tgctgagcga aacacgcttc actgcaagat gctctcatcttc 660
aaaaattctc cggctgagac gtaattttttgt gcgtgagcgc ccagccgggtct ccagacttac 720
ttcagcaagt ttcgtgtgat gtgcgggga aagcatacct ggtgttaacg tgtgagattc 780
gcacatcatc caacctgcgcg acacagcaact cagccatcttc tgtagcacc ggcaagattgct 840
-continued

gactacttcc atgcctcagc tgacctaaat acctgcggcg ctgcctccc tccatgtgctc 900
cagaagcaca gtggaggtgc acgtgccgta ccggatccgc ctgcctctcg 960
cagatgactc ctcagcagcc ttccggggaat cctgaaata tattgctgaaaa 1020
cggagagaga cggagatttgc tggatcagctg gaaatgtgcag gagaatggag 1090
cgagaagggac acgtggtgct acagctgcg gcctcccaga tccagatgtc 1140
aacaatcagc ttatggtctta gcgaaagcct acagcctgcca cctacgtgac 1200
cctgcctgctc gctggccaaaa aaaccatctt cccaagtcaat accagagttg 1260
catactcgcc cagaacagca aagttgatct cggcaataag ggcgctgag 1320
gatccgatt gacgagcgtg gccatgggtt caccggatggc tgaagctttg 1380
aacacgctg cacaagacct cggaaagttgc cccatgcgct gttcgtgctc 1440
agtctgctaa aagccgaagtc gctgggggtg agcagatgatt ggttgcttac 1500
cggtctcagc atgcgtgctgg cctgtgctcc gcctgtgctc atggatgtgat 1560
cgataagga gacaaggtga atcgccgaac gcctccgaa atacccctaga gag 1620
tcgcctaac gcctgagcc caagggctag cccggatccgc tttcgctcag 1680
aaccattctt taccacgctc tctaataaaa cagatactcg aatcggcgtg 1740
gagggaaaaa cattcattcc tccctgctgt cccatgtgct caccagcccg 1800
tgaaacgcg tggaggtcag gcgtgagcct gaaagcgctc cccttggggc 1860
agacataact ctgccccgca gaaagtcgac tcttggtacttg tgcagccctg 1920
actgagcacc caccctccgc gctctcggct gaccccttt aacaagcactc 1980
acggcagcac gccatgtcctg agttacggtat cccatgcgct gttcgctgctc 2040
tcgcctaac gcatttttgc cccagcagc gcctctgctgct cccctgcctg 2100
cgagggcttc tgcagcagcc gcgttaccc aacaagacca cggggtggca 2160
gctatctgg gcgggagtcg ctgtcctcgc cgggcggctg tcttggtctgtc 2220
tactttaata tccctggtgg cccagcagc cggcggctg tcttggtctgtc 2280
aaattttcag aatacgtgcg gcaccaagcg cgcagcgaga tgggctgctgggc 2340
cgatctgagc cactcatcgc cccctttgct tcaaatgatgc gaaagctgtct 2400
gaattctgcg gcatttttgc tgggggagtc cggcggctg tcttggtctgtc 2460
actacccgag ctctctgctc cggcggctg tcttggtctgtc cggggtggca 2520
catactgact ttcctggtgc aagttgatcc gcctctgctc cggcggctg 2580
gagggcttc tgcagcagcc gcgttaccc aaaaagctgct cggcggctg 2640
tgggctgctg gcatttttgc cccagcagc gcctctgctc cggcggctg 2700
gacggcagc gcgggggagtc gcgggggagtc gcgggggagtc gcgggggagtc 2760
actagctgcg ccctgtgctc gctgcggcg ggcgggaga ctcctctctgctc 2820
tcctgtgctc gcgggggagtc gcgggggagtc gcgggggagtc gcgggggagtc 2880
aataccgccga caggagctgc gcggggagtc gcgggggagtc gcgggggagtc 2940
tgggctgctg tcctctcctc cggcggctg cggcggctg cggcggctg 3000
agaagttggc gcgggggagtc gcgggggagtc gcgggggagtc gcgggggagtc 3060
cgagggcag gcctctgctc cggcggctg cggcggctg cggcggctg 3120
-continued

cactctcagt ctcggtcagc tctggctctc caagctgggg tgggtgcacg 3180
aaccgccct gtacggtcag cgcggggtc agaggttcgtg cccaccaacc 3240
cggtaagaca cagactatcgc ccacgtggag cgccacacttg tgaacaggatt acagagcagca 3300
ggtatgtag ccaggtgtcgc ctgcttctgg aatcaggttac gcaagacacca 3360
ggacagttg ttcgtatgct gccttcgtgta aggctagtttc cttcggaaaa agagtggtgta 3420
gcctttggtc gcggacacga accacgcgtc gtacccggttg ttttcttgat tgcggagacg 3480
agattacgag cagaaattgg gatctccaag aagatctcttt gatctttcttc aaggggtctg 3540
aocgctactgg gaacagcacc tcaagtttaag ggtatgcttg catgagcttt acaaaagaga 3600
ttcacacatt gatccttttt aattttatttta gaataaagat aacatttatc taa 3660
agtttaaccttg gttgactcagt taaccatgctg taacactgtgag ggcacacttt ctcagctatc 3720
gttacttctg ttctactcatt gttgcttgac tccocgtcgt gtgataaact aagatagggg 3780
aggygctacc atcctggcccct agctggtcagc tgattcccacag gccacagcgg tcaagccgtc 3840
cagtttattt acagactaattc cagcacacgg gcagggcgag aacaggcaaggt ggtacctcga 3900
cattttacgg ttcctactcag tccttaaattt gttccggtag aatcaggttttt aagatgcttcg 3960
cagattttag tttggcacaag gtttcttgcag catggctgggt gatactgctg gtaaactgtg 4020
cgtgtgtcac gttctctccag cccacagacgc aacggctagttt ccatgatcctc 4080
ccaggttggc cagaaacctg gattagcttt cggccgcttc tggctcccagc cagagctcagc 4140
ttgccggcaggt gttacactgc atggatctag tagcagctgca taattctcctt aatctgctagc 4200
cctgctgagaa atgtggcttt gttgactggtg agtctatccg caagactttt tcgactattt 4260
gactgggacccacaggtgg cttcctccag cgggcttaag ggtatatcacc ggcaccacacta 4320
gcggcaggg cagagcggttc atcctttggg aagctttttt ggcggcggaaa cttcttgcag 4380
ttcacactggt ttctgagacag agtggctattt acagactaattc tgcagaaattt tgcagatcag 4440
caatcttttc ttcgcacgag gttctgcaggt gaggaaaaag aagaaacccg aactagccaa 4500
aacttgaactt atcgagtcag cggagatagtt cagcagcagcg gcttttccag gcaggttggct 4560
attagatag ttcagaggttt tatggttctca tcggcgtact caatattttc tgttttgtcg 4620
aaaataaata cattggtgct ccggcagcat tactccgaaaa aggctcagct gacgttcataa 4680
aataactatt tatcagcagcctttactata aatattaggg tagtacagcag ccccttttcag 4740

<i>2<sup>10</sup></i> SEQ ID NO 152
<i>2<sup>11</sup></i> LENGTH: 4746
<i>2<sup>12</sup></i> TYPE: DNA
<i>2<sup>13</sup></i> ORGANISM: E. coli
<i>4<sup>00</sup></i> SEQUENCE: 152

togcggtttgt ggtgtgatagac ggtgaaaaacc ttgacacat gcaagctcgcg gagaaggtcag 60
cagcttggtc gccggtgccgtg gacacacgca gacagggcgg gcacggcgcg tcaccgggggt 120
ttgcggggtgt gcggagcgtgct gttactatg gggctcagag cggagattta ctgagaggtgc 180
accatgagcg ggtgtggtat ttcgtgagat cggcagtttta cggagagtcg gcacggcgcg 240
attgccgctt cagctggtgcac aactgttttgag aagggctcagag cgtgccgggct ttctgctat 300
tacggcagct ggccggaaagg ggtgtgcttg cagcagagatt aagttgggta aagcagaggtt 360
ttttccgtctg cagcagctttg aaaaagccg gccggtcgcct gtttaattaa tttcttgctg 420
gaatcggccca agcgcggggg agaggggtt tcggtattgg gcgcctttcc gctcctgcgc
2760
tcactgacgt ctgctgctgt tgcgtctggc tcggtcgag cgatcagctg ccactaaagtg
2820
cgtaatacgt ttattcccac gcgtcaggg ataacgcaag aagaacactg tgagcctaag
2880
gccgcagccaa ggccgacagt gctttaaagg cgcgcgttgc gcgggtctttt catagggtcca
2940
gccccctgca cgagcactac aaaaaacagc gccttaaggtc gaggctggca aaccggcaacg
3000
gactataag atacccagcg tttccccctg ggacgtctct cgtgcgctct cctgcttccag
3060
cctgcgcgt taccgattac ctggtcgcct ttcteccttt gggaagcgtg gcgcctttctc
3120
atcagcctaag cttcagtgat cctcgctggc tcggtgcttg tcggtcagaa ctgctggcttg
3180
tcgacgaaac cccccctcgc cccgccgcgg gcggctttt tcgggtactt ctcggtttgaat
3240
ccacccggtt aagacacgag cttctgcacac tcggaacgcg cactgtaaac aagattacgc
3300
gagcgaggtta tggaggggt gcgtcagagtt tcctggaagtgt gttggcctaac taccgctaca
3360
tcagagacgc aatggttggct atctgcgctc tcgtaagcgc accctctccc ggaaacagac
3420
ttggtcgtc tgtacgccgc aacaacaaaca cccggtgttac ggccggcttt ttgggttcga
3480
agcgcacgat atgcgcgcaac aaaaaaggct ctcaagcagct tctcctgtagc tttttcctcg
3540
ggtctgacgc tcgacggaac gaaaaatcac gttcaggggt tttgctctat gcattatatca
3600
aaagatcttt ccacactatg ttttttaaat aaaaatagtt cttaaatcta atcctaagta
3660	tatatgata aacctttgct gcacdgctcag aaatgtaaat aagtggaagca cctatccatcag
3720
cgtatgtct attttgttca tctacgcttc gcgggtcgtg ctcagtaacta gcacgtcga
3780	tacgagggg aatccactgt gcggcctaggct cgcaatgat agccgcagcag ccagctcaac
3840
cgtcctccagc ttttctacgca ataaaccagc cacgcgggaag gcgcagagcgc agaattggctc
3900
cctcctccatt ctccgctcact accggaatct tattaggttt gagggaggct aagatgaagct
3960
gtctgycagtt taataggctg ccgcatttgg tggccactgc tcacgctaagg tcgggcttcag
4020
gtcgcgttt tgtgtatgtct tcatccagct cccgttccga cagctatcaag ccagctatcag
4080
gatcccccct ttttggccaa aaaaaaggtta gcgccttggcg ttcgtccgatt gcgtcgaaac
4140
gtaagtgcc gcgcgggttga tcacgtcttg tttggtccagc acctgcataat tcttttcctg
4200
tcgtgcaactct gtaaagttcg tttttttgtc gttcgtgatt ctcaccaaaaaa tctcattgag
4260	aatgtgtat ggcggcagcc agttgcttct gcgcgcgcgcc aatacgagat aataccggcc
4320
cacatacgg aacttttaaaa gttgcctcacta tttggaaacgct ttcctgcgggg cgaaaaacttc
4380
cagggatctct cccgctgggt acgtgtgtaa ccagctgcgtg ccacactcgtg
4440
cctcggcagct ctttaccttc gcgcgcttgtt tggggtcgac aaaaaacggca gggcggcttg
4500
cggcagaaaa gggaataagg gcgcacggca aattgtgcta aacttacactt cttctttttttc
4560
dattatgtt aagcatttttc agggtttttt gcagagttag gcgatacaaa ttttagctgtca
4620	tttgaanaa tansacatata gggttctcgc gcacatttcg ccaaaagtgc cccactgtgac
4680	tcctcagagcc ccattttata atacgacattaa cctttaaaaaa taggctgatac gcggggcctt
4740
<210> SEQ ID NO: 153
<211> LENGTH: 3946
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 153

tocgcggtgtt ggtatgagct ggtgaaacc tctgacacat gcagctcgcg gagaoggtca 60
cacgttgctt gttaagcggat gcggggagca gacaagcccc tcagggcgac tcacggttcg 120
tggcggggtg tggcggcttg attaaatag gcggatcga gcagatgtag ctgagactgc 190
acacatatcg gcgtgaaata cacggacagat ggcgaggag aaaaatcgc gtcgacgcgcc 240
attgcgcaatt cgagctgcgac aacctgtagg aagggcgact gtcgaggggcct tctcgctat 300
tacgcacgtt ggccccaggg ggtatgtcgt ccaagcgatt acgtttggtta aegccagggt 360
ttttcagacg acaacgtcttg aaaaagcagg ccaagcttcc caagtcagcc gtcatttgct 420
tgatcgctga caaaatagta caaaatgagct ctcacttttc ttcacaacgc 480
gcaaggaact cggtcgccggt ggccccggtg cattttttaa taaccgcgca gaaatagtg 540
tgatcgctca aaaccacatt gcgg accgaag atgtgcagat gcaagtcgggt ggtgcctaaa 600
agccgtcgct tgtccgtctg atcgatttgc tcgcgccccg tcctcgactcy ctactccctac 660
tgcggcggga aagagttgaa cagacgccgac gcggacaccg aacaagactg cgcagctggc 720
gcagatatgg tgtccagcgtt gcggcctgtc gtctcgagctacctccgc caccgccgct 780
cagattcacc cccgaggta ccaattcgga ggtggctgct ctggtggtatc gcgggtgac 840
tgtccacgca ccattaacgt ccagcagggt ctcccccctt ccggcctgta 900
atgatttgc aacaagcggt gcggacagg gcggcctgct ctcctgagct gcggacaccg 960
cctggttag ggccatgctg agcccggtta accggctcct gcaagtgccg ggcagcaccg 1020
aagtaaaccc accgtgggtc acaattcgga ggttcgggat cggagccac gttgactgatc 1080
ttctcctggg gcggacaccg aaaaatccg aatattccac gtcggccaaa caaattccct ttcgtgattt 1140
tttcgcaccc ccgccagcgcg aagttggagc tttgagatttt acocctcctcc tcccgccggg 1200
cgggtgaata aaaaaatcag ataacggcgtg gcgtcatacg gcggtaaacg cggccagcga 1260
tgggcaatg ccaagtatcc gcggacaggg gcggcctgct cttccggtcc gcaatttctcc 1320
ataagcccgg cttgcagcag caaaccacat gttcttattt gtcagacgcc ttggtgcctcc 1380
tgcggttcttt accgtgtcct tgtacctacc accggggtta cccccctgatt taaaaggctt 1440
tgttaaaaac gggccaggg accggcagtt gggagaggtt aacaagttg 1500
gcgccagag ttcattgctg ttttttcgcc gcgggtccac ttcggatagc ttcagatttt 1560
ttatccattt gattttcggg cttactcctg cggcgaactt ttcggttttt 1620
cctactgctt ttttttttttt ggtctcaggg gcgtgcgacct gggtggtagc cggccctgctc 1680
atcggacaag gcggagggg ttcggttgcgc atcctgcaag atctgtccata gttggttccc 1740
gtgatagcc ggttggacat accggctgat acgtggcata ccagtcttct 1800
aagctgggg ggtgactcag aagctggg cttccatgtaa tcggctgcac tgtctgctgc 1860
cttcctccag cggggagact gtcgggacag cttactgtct acgcggggg 1920
agagggctgt gcgcctctct gcggctgctg tcgtctgcgc gtcgctgtcg 1980
gtcgcgctg gcggcgcgag ggtttccagc cactcaaaaag cggtatagac gttcattttc 2040
acatacgagg aacagacatg tggcagacag gcgcccaaag gcgcagggac 2100
cgtaaaagg cggggctgttc gcgggtttttt ctaggctcc gcggcctgtg cggcgatcac 2160
aaaaatgac gtccaggtca ggcgtgcggag aaccgcagct gactatataag atggcagcgg 2220
tttccccctg gaaagctcct ctgtgcgctct cctgttccga cecotgcgct taccggtacc 2280
cagcgtgct ctctcottcct tggagagaagtg gtttctctctg atacagcagct ctgtaggttat 2340
tctagtaagc tggagctgctg tctgctcaagct cttggctgtg gcaagcagc ACCctgttcag 2400
ccgcacgtg gcgcctctct acgaacctac ggttttggtg caaaccgcctt acagaacgac 2460
ttacgaccac cgccagcagc cactgtaaac agaatagca gcagcgagga tggatgaggt 2520
gctcaagctt gttgacagct ctaagcctatactgagcat ctaagggac agtagttggtttg 2580
atacgctgct tggagttaagc gcgttctccttc gaaaaagaggt tggagctgct tggatgagggc 2640
aacaacaagct cctgcttgat gctgggttttt tttgtagtcac gcgcagcagc tacagcagca 2700
aaaaaaggtt ctaaagaaga ctcttttagg tttttctcgg ggtctgaagc tcagttgaaaagc 2760
aacaactcacc tttggctttttg tagttatcag aagatttataa aagaattctt cccttagaacc 2820
cattaaataaa gaaagaattt tttttatcatca atctataagta tatataagta aaccttggattt 2880
gccggatgct cagagcagcgc cagatctgtt gtttgtgtttt tagtgtggta 2940
tccttgtccgct ctgctgttggtag aaaaaatca taacagcaggtt gttgtcaattc 3000
gccccctcagct gctgtcagagc cagccaggaacc cccgcccgctg cggcttgcag ttttacgagc 3060
atataaaccac cccggaaggcc gcgcggcagcgc ctgccccatct ggcagctgtcc 3120
ctattcgacta taatgatcgcagc cctgtcctgtc ggagagctcagatatggtagatc sgatgccg 3180
tcattcgacta taatgatcgcagc cctgtcctgtc ggagagctcagatatggtagatc sgatgccg 3240
tccttcgagtc acgtgttagc gcgtgtgtgg gaatttggttg cggccagctg 3300
aagctggtta gcgtttcttgctctcgagttctgtggagttcagttgccagcggccgttgg 3360
tcattcgacta taatgatcgcagc cctgtcctgtc ggagagctcagatatggtagatc sgatgccg 3420
tttgtctgca ctgctgggtagct ctaacccagc ttgttgctagat tctaggttgtga ggggcagcc 3480
agtgttcgct gcggcctgtc aataaggagc aataccgagtc ccatgccagc aacccctaaa 3540
tcttgctctac taagcagccctct ttttctctgctg gctgttagct cagccgtttg 3600
agatcagctgt gacggtcactg cctgctgagc ccccagcctgt cttccagatt ttttacacctc 3660
acggattctg cttgccagt gaaaccacag ggcagcagcgc cgcacacgggg ggggaattgg 3720
nggcaagat gatgttggattc atctgaatc tcccttttttt taaatatttt gaaatattatc 3780
cgggttagt cttgcttgagccagcatactg tttgaggtgta tatgaaaaaa taacccatag 3840
ggggtttccggc gcacatcatcc cgcacagtcgt cocaagtcag tcataagacac ctttatttttc 3900
atgcacatata ccctatattttu tgggagctgt ctcgct 3946
-continued

taccggagct ggcagaaaggg ggatgtgctg caagggcgatt aagtgctggta aegccagggg 360
	ttcccagctg acagcgtgct aaaaagaggc ccagtgccaa gttcgaaggc gtcacaatgtc 420
tgatctgata caaatattgac caaacttgag gccacactcgg tcaacttttc ttccacaacgg 480
gcagggaaact gcctgcagct gcgccccgtt cattttttta ataccggcgaa gaaataggt 540
tgatgtcgaac aaccaacact ggacagccag gttgctgata gcatccgggttt gtgtgtccaaa 600
tagccgcttg ctcgacctgat acgtgctgct tcgcgcgcgg ttgagggcctt aaaccctaatc 660
tgcgtgggga aaaaaaaaaa caacgcgggac gcggccagccaa aacatgctg cggcacgctg 720
gccattactc aatagctgtt tcgcaagttg aacactgtgt actgacacag ctcgcgatct 780
cgatattccca tcggtgtcgg gaagacactg ttaatcgctt ccattgcgcgg cagtaaacaat 840
tgctcagacc gattatcgcgc cagcagctcc gcattaagcccc cttccctccg cccyggcgtta 900
agataggcccc ccaacagctg gctgaaatgc ggtgctgcgg gttcatccggg gcggaaagac 960
ccggttagtgg caaatatagga cggccggtta agccattact ggcagatgag gccggcagccga 1020
aaggtaaccc acggtgcgtga cacttcggca gcttacgctg gaggacgctgt ggtatgaatcc 1080
tcctctggg gcgaacgcaaa atacactccc gctggcccag ccaaatctcg tcctgatatt 1140
ttcacaaccc cctgacaaggg aagttgctgaga ttgagaaatt atacctttat acctgagctt 1200
cggctcggagg aaaaaatctag ataacccgcgt gcctcatacg gcctaaaccc cgccacacaga 1260
tggccattaa acagttacgc cggccagggg ggtaatttttt gcggcgtcagc catactttttt 1320
atatcggcg ctcagccaga agaaccataatt gcgtcataatt gcataaggacac tgcggctcgc 1380
tgcgccttctt tgcctgctttt ctctgttacaac aacccggtaa cccggttata taaaagcatt 1440
tctgataaaaa gcggacacca agccatgacca aaaaaatgtt ataatattttt cctaataa 1500
ggcagaaggg tcccacattgc tttttgggcac gcggcgcgcac tttttatgtgc atagcatttt 1560
ttatcattaatt gattacggtt cctctacgctg gcgttttttt caacatcore ttcggtttttt 1620
catacgcgtt tttttggggcc ttgaggaggg cctgtcaggt gcacatcaga gaggc Georgetown 1680
gccctcaac gaaagggcttg aagggctggtc gctgtctgtttt atgttacctt gagaatgtttt 1740
titcttgatt gcattggttta tccgtcagatg ttcacacgaga acataagcagc gcggagctca 1800
aagttgaaac ccggggttggcct ctaatggttg aagttactacc ctagatttttt tcagcgtca 1860
ctgcctgcct tccgtccgag aacactgtcg tcagagcctgc attaacttgaat ccggccagnac 1920
gcggaggagag ggggggtttgc tattgggctgc ttcgcatgttt cctctgttgc cagagcccgc 1980
cgctgccgct gctcatttcgc gcgaccgccat ggtgggctga atcactgcct 2040
tccagcccc cgggggtaaa cgcagggaga cgctgtgaga acaagggggc cggaaaagggc 2100
aggacgctct aaaaaatttgc gcggagggcg ttcgctggggg ttcttccatga gcggcgcgcgc ccgccgaaggg 2160
ccacaaaaaa atcaagctgc cggccgaaccc ccgacagcaat aaaaatgata ctcgagag 2220
caggggtttcc ctcctgggaaag ctcctgcctgt gcgtctttcg tcggaggtc atccggtcacc 2280
ggatgctgtgg cgcgcttttct cctctgggag acgtgcgccc tcggctcataag ctgcgtctgttt 2340
aggtactctc gcgtattggtc ggtgctgctgtt tccaagctgtgc gggtgtctca gcgaaccccc 2400
gtccagccgc cccgccgctg gttcagccct atctacgctg tgggtcttcaaa ccggtgtaaga 2460
ccagactttc ccgcagctgc gcgcacggacct gttacacgggct ttagcgagag gaggcatgta 2520
ggcggggtca cagagtgtttt gasgtttggct ctaaactacct gcgtcactag aagggagcta 2580
-continued

tttgtactt ggctctggct gaaagccagt ttacctggaa aagatgtaag tagctcttgga 2640
tcggccaaac caaaccgcctt gcagctgcttg gttttttttt ttggaacgct gcagattaacg 2700
cgcggagaaa agaagaccttc ttgtgatttt ctacaaggggc ctacgcgtagc 2760
tgaacgaaac ccacacttga aaggtatttt gctcaaatag tataaaaggatgctacc 2820
tagatctctt taatatataaa tataaatataa caagataact aatgtataa tgaagtaacc 2880
tgctctgacac ttaaacatgt ggccacacaa aatgtagttt tgtgatcttg cgcgtctttttt 2940
cgtctcatcc tagtggcttg actcctggctc gtgtgatataa ctacgataac ggagggttta 3000
cactcgccg ccaagtgcttg aatgataacg ccagacccca gctcaccgcctc tccagattta 3060	
tccagcaattt accagccggc cggggagggc gcggcgcgaa gttgtctgctgc aacctttaac 3120
gccctctcc agctttatcca tgtgcacggc gaagtgactag tgaagtgcct gccgatttaaat 3180
gattgtgcca aggtggttgc cactgtcaaa ggcatacggg tgcacagctc tgtgggtggt 3240
atggttctct tcggttcggtg ttcgccacga tcgacggcg ttcacctgatg ccctcaatg 3300
tcgcaatagc ccggtgtgcttg tggggaatcgc cggatagttaa aaggtggcgcg 3360
gttatcactc ccctgtcatc gcgcaggcttt catatccttc ttaactctctg gcctcatgct 3420
agatgtctttt ctgtagcttg aagqtgccttc accagcttcat cttgagaata gttgtgtggg 3480
cgaccggagt gtctctgtgcc gcgccctaaa cggatgataa cggggacca ctcgacgaact 3540
ctaaaaggtgcc ctctctcttt cggagtcggttaa aacggtcggg gacgtcttacg 3600
catgtgtaac gcattgctttt gttgttcgctt gcacgacacca gctcagttttt gcctcaatctt 3660
acccctccaa gcgtttcggg gttgagcaaaa acaaggaagc cagactgcgc aaaaaaggg 3720
ataaggggca caagagaaat ttgaactact atactcttttc ctctctctctcta ttttctgaac 3780
attttcttgg cattgagcgg cgccttctatg atgtatcttaa aaaaattaat 3840
cattagggg tctccgggca atctcggccga aagttggccac ctcgactctg agaaacactt 3900
attttcttgg catttacca ctaaaatagg cgtatataac ggccctcttcg ctcggttaaa 3962

<210> SEQ ID NO: 155
<211> LENGTH: 3896
<212> TYPE: DNA
<213> ORGANISM: E. coli
<400> SEQUENCE: 155

tcgccgctttt ccggtgatgc ggtgaaaccctctgacacagt gcagctcgcgg gagaaggtta 60
cagctgtgtcc tgaagcggtt gctcgggacag ccagggcgcctcgc gcaagggcgttttttct 120
ttgagggggt cttgcacgattt gctcgcgatatg agctgcgtgctctg 180
acacatgctggagtgatataa ccaaccgacg atatcagagagg caccagctcagatcctg 240
attgcgagtt gcgcggtggat gcgtgaggtttaa agggcgtcattgagggccgtc cggcagttatct 300
tacgcaagt ggcaggggctt ccagagttttg agttttaaatgcagttggtttg ggcagtggctt 360
ccccggctgg ccagctgttttg aacaagccggcc gctcttacgcttg ccctcctctctcctctt 420
tgagcgcggcgcctgtagcgtctgcttgcctctc gctacactccctctcttggcggccgagcacta 480
aagtggttgagttatcctgcttaagctcggcctcctcgagcagcctactcttgacctgacatgggtctt 540
tcttttcatgg tagctgcaac cagctgtctggtgctcgaag ggcaagttttttttctcgcgcgcttct 600
tgctttcaagctcgtctgaatgcgcatggctttgctgagtgctctccttgagtttattttttctttct 660
-continued

ggcaccaggtgtgcaaatggccagatccgcagacgaccgcctacccgggtctccgaagttatcgcacaacgttatc
gttatcgcacccgggtctccgaagttatcgcacaacgcctaagatttatggccatgggttattc
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}
ttcttatttttattctatcctatcctaccaacggtctccgaacgtatcgtacgttttatttatttttttatt}

<210> SEQ ID NO 156
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 156

gcagactacctgctataatttctaaagtac 28

<210> SEQ ID NO 157
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 157

cgcgggtaccaaatcgccctttgtgattagggcgcttttggcgcttttcaggtctgacgtcgtc 60
acaaaccctcgaatttcaagttggc 86

<210> SEQ ID NO 158
<211> LENGTH: 86
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 158

cgcgggtaccaaatcgccctttgtgattagggcgcttttggcgcttttcaggtctgacgtcgtc 60
agaaaccctcgaatttcaagttggc 86

<210> SEQ ID NO 159
<211> LENGTH: 38
<212> TYPE: DNA

<400> SEQUENCE: 159

cgcgggtaccaaatcgccctttgtgattagggcgcttttggcgcttttcaggtctgacgtcgtc 60
agaaaccctcgaatttcaagttggc 86
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 159

tgcagggccc tctgtgtagc ccaaaaaac gggatgg 38

<210> SEQ ID NO 160
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 160
cgcgcgggtaa caatacgc cc ttggtgatga gggcgcgggg gctctctgaga gtcgacgtcg 60
agggcctcct gcctgcacccaa aaaaacgggt atgg 94

<210> SEQ ID NO 161
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 161
cgcgcgggtaa caatacgc cc ttggtgatga gggcgcgggg gctctctgaga gtcgacgtcg 60
agggcctcct gcctgcacccaa aaaaacgggt atgg 94

<210> SEQ ID NO 162
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 162
cctcttgcggt tcggatata accttcagct gg 32

<210> SEQ ID NO 163
<211> LENGTH: 91
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 163
cgcgcgggtaa caatacgc cc ttggtgatga gggcgcgggg gctctctgaga gtcgaccccc 60
tctggtctgc cttgataaac ttcctcaggg c 91

<210> SEQ ID NO 164
<400> SEQUENCE: 164
000

<210> SEQ ID NO 165
<400> SEQUENCE: 165
000
---continued---

<210> SEQ ID NO 166
<211> LENGTH: 2319
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 166

gtgcacatga aataaaaaac aggtgcaagc atctcgcatt tatacgctt aacgacagtg  60
atgttctcgg ccgcgtctct gcgaaaaaat gcagaaggtg aacgtgtaat ctggattacc 120
ggcgttaaga gctataaacc tctctgctaa gttgtaaga aattcagagaa agatacgcga 180
attaaagctc cggttgagca tcggattataa cgaaagagaga aattcgcaca ggtrgocgc 240
acgctgccgt gccgtcacta tatctctgg gcacagccag gctgtggtag ctacgcctaa 300
tctgcgcgttg tgcctgaaat cacccgacac aacgagtttc aagacacacgt gttgccgttt 360
acctgggagat cggcagcgtta caacgcggac ccgccgctcg ctggtggtagct gttgcaagc 420
ttatgcgta ttttataaaaag ggctgtgctg cgcagcccgc caaaaaaacctg ggagaagatcg 480
cggpccggtg atataagaact gcagaagaaaa ggtaagagcac gctctgattgg caaacctgaa 540
gaaagctact tcaacctgccc gcgtgatttg gtcggtgaggct ttatgtggct caagtatgaa 600
aaccgcgaag acagacattaa aacgtggtgcg tcggatagcc cgccagacgg aacgcgtgtc 660
aacctcctgg tgcgctcttg taacaaacaa cacacaagtac gcacacgcg ttaacctgcac 720
gcagacggtgt ccctcttaaaa aggccagaaa gctggatccga taacagccgcgt gttgcaagtc 780
tccacacagc acacacggaa agtagaatatt ggtgtacccc tacctgcagc ctttaaggtc 840
caacacacca aacgcttetg tgggtgcgctg aacgcaagta taacggcagg cagccggac 900
aaagacgtcg ccggagaaaatt ctcctgctag tctctaggaac tctgcgagct 960
gtattataac ccacaaacgct ggctgctcgtg gcttgagat cttcggagag aagaggttcgg 1020
aagatccac gcattggcgc caacatggaa aacgccggaa aaggtgagat cattggcgac 1080
atccgccgaa atgtgctgtc tctggtgatcc gttgatggtg aaccccgcag ctcggatcgc 1140
gacgctgacgg ggcgcaacac gcacctggcag gcgaacactg acatttattac caagygtcgtg 1200
gttgaatctag tatacggtg ccctcctgcct gttggaacct cggagctaag cggatgagc 1260
ttcctcctgg ccggtcgaaac gcacactgcag ccgattggccgc acgatgctgcgcc ttcacccttg 1320
ggcgacgtcg ccagtgcgct gccggttatct ctcgtgctgg ccgtcgtgcgt ggagcataac 1380
aatcttcctg gggttcacca tctctgcggc ttgggtgagct cttcgctgctg tgtgtactat 1440
ttcctcctgg atgcgtcggc ctattccacc gcaccaaatag taagccagct gatcgtcgag 1500
cgtacgctg ccagatgcga tctctgcggc tctctgctggt cttctgctac ggtgtgagg 1560
aaggaaactca taagctgcac atgctgtacc ttcgctgccgt ggtgtataac cttctccttc 1620
aacctgtgacg gcctgacgcag cgtgtgctgca ggctgtcgcctt cttgctgctg ccacctggtc 1680
cccattgctt acacacgcag tcgtcaggtcc gtcattccggg ttaaccaccc catgtccttc 1740
tctggtcac ccctctctata aacacgctga tggcaccacaa acgtgccacgtg 1800
atgccgtcgg gccggcccgcc gcgggtgctg cgggtggccg tgggacacac caacggtttaa 1860
ggcgacagc cgtcagcagc gcacaggtgatt ctggggtgcgg ctcacccacct cagccagcgcag 1920
gtccgctgtc tagcgtgctg gttcattggc cttggtggtcg cttcgtgctg 1980
-continued

cgacgcctga tggctgtgta tattcttaat gaaacagtga gtaacgtcct cttggtttcc 2040
ttcctggtag gtcctctag acaacagctgc cttcttcatg taagctcgcg ctaaaatctc 2100
atctctctac acctggttctc cgcaaacattc gcaacggtct ctctgctcaac gctgggctgc 2160
cctgggtctag gcggcggcgg aacgacgaaag aagggcggac gtttctgcag gaaacgccaa 2220
agcagtgtact ccgacccactt ctttcacc acgctgcccaag gggagacccct gtaacgctgc 2280
gcgattata aagacgcgta tgcacaaatct atggttaacc 2319

<210> SEQ ID NO 167
<211> LENGTH: 1293
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 167

gtgcagcata ccataaaccg cattgcacgc atctccgcat tattctgaat aacacgcatt 60
tagttttcag cttgtgtgct cggcaaaaatc atcgaagccg cgaacctgcgg aacgacaacg 120
gcggcgcaccc acacgcaccc gcacggtgac atcagccatt attoacaggt gctggtgact 180
gtcataactc tggcagcgtgac gctggggcga actcggtgag acctgcttct 240
tggaggcgcag gaaagccactc gcaaacgtct gcaagctctg tcgtagatc aacccgacgc 300
tggcagcctgc gcacgactgt tattctcttg ctgagcctgc gcggcggagct atcacaactcc 360
ttcctggact acataccccgt gccttttggg gcgcgtgtct gcggcgtgta ctattccttcg 420
cgctgtggtct gcacatcact ccacacccaat gaggagcttct gcggcgtac 480
tggccatcc gcatacctgc cagcagccag acctccattg ccaggccgcgc ccaagaaaa 540
ttcacgctg ccatacctgc gcggcggcgc cggctgacgcg ccggtgctgc ttgccttactc 600
ggcctgcag gcacgcagttg tgaacgcgg cccctcggtgg gcctgtgttg gcacacccatt 660
ttgagccacac cgcgctcctc gcctcgctct aacagtcgtc aacctccagct toctcttcttt 720
cctttgggtct cattctcaca ccataaccct ggaattgcag ccacacagttg gcgcggcctg 780
cactcgcgg ccgacgaccc ccagtttgct gcactccggg cccacacagc ttgctgccac 840
agaagttcgg cgtcgcctgc gacgtgtgcgc gctgcgctgc gctctcggtgg ctggggtgcag 900
gcttacgtgc gtcggtcgct gcgtgtctgc gcctctacca cgcggcggac gcggcgtactg 960
gtctttgcgc gtagcagcct gctggtctgc gctgtgttcgt gcgtgtgttg gcacggagtc 1020
ttcctctgac aacaaggcag aacctgcagcc gtctgcagct gcacgtgggg gcagcgcaggg 1080
tctctcctgc gcccagcagg ccagttgctgc gctctcgcgg gacttttcgg cagcagcagctg 1140
tctcggtggcc gccgagcgcc aacagctggg cccagtcgct gcagggcggc ccacacagctg 1200
tctcgcggacc ggctggtgcc gctctgctgg gcagacgggc aacaagagct gtctgctgcgc 1260
ttacaagatg acatgtttgg cccacacgcc acggcggaga tgcctgacgc gcggcgcagat 1293
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 169

gtgcacatga aatataaacg aggtgcacgc atcttcgcac tatcgcatt aacgcgtatg  60
atgttcccgt ctctggtctct ggcgsaactc gaagaagctg aactggttaat ctggattaac 120
gggtataaa gcgtataacg tcggtctgaa gctgtgtaaga aaattcagaa agataccggg 180
atataaggca cgggtgacgca tcgggatcag ttcggggtat gattctggcc 240
actgtgtcgg cctcgcctgat ttacctctgtc gcgcacacgc gctggtgctt ctcgctccac 300
ttctggggttc gggttctgatg aaacctacca aaagcgtgcc aaagcgtgcc ggatccagtac 360
aacgccggtgc ctcagttcct aacggtcctgtg atgatcgttc accgggttgc ggatccagtac 420
aacggttcgc ggttaacaag ggtgattcgc gctggtggtgc ggatccagtac ctcggttac 480
gtcgcggcgttt ctcggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ctcggttac 540
gctggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 600
ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 660
aacggttgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 720
ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 780
gtcggtgtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 840
gtcggtgtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 900
gtcggtgtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 960
gtcggtgtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1020
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1080
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1140
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1200
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1260
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1320
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1380
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1440
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1500
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1560
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1620
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1680
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1740
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1800
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1860
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1920
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 1980
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 2040
aggtgccggg tacacggtgg ggcgggaggc ggtggtggtgc ggtggtggtgc ggtggtggtgc 2100
atctcttca acctggtctc cgccaaacct cgcagagttt ttgtgccca gctgggctgc 2160
ccttgctctg gggtgaccca cggccagaag aagagcccaat cgttttcgaag gaagccccag 2220
agcatgcca gcacccatct cttttccacc aggcocaccg ggagccctc gcaccgggcc 2280
gcagcagata ccaatatttcca ccctgacctg acgatgttgg acagcagatt acctaaacgcg 2340
ggagggggca ttcttccgaa tttttgcca gaaggttgtgg gttcgggcaaa atgtactgcc 2400
cctgttctgg atgaatccgc tgcagattat caggccaaac tgcagcttgc aacaatgcac 2460
atgcattacaa aacctggtcgc tcgcgcgaata tatggaatcc ctggtatccac gcctctggtg 2520
cggcataaaa agcaggtcag ggcggcacaac aagttggtgt cactgctcaa agtcaggttg 2580
aaaagatttc ccgacgctaa cccagggggg ggcgcagatatt aataagatga cggtagccaa 2640
taataaggtga cc 2652

<210> SEQ ID NO 170
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 170

gtgcagatga aaatcttccac agggtgcacgc atctctgcat tatcggcttt aacgacagatg 60
atgtttcccg cttctggtcct gcgcaaaatc atcgagaccc gcacccctggg atcgcacaag 120
ggagggca gcacccacat ggaagttgga acactagcttt atccaaagg gttggtgtac 180
gctatatacc cggcgcaatt ctggctgggc ctgtctttcg aacctggttac gcctctggtc 240
cgctgctgcag ggcaggtcag ggcggcacaac aagttggtgt cactgctcaa agtcaggttg 300
ggtgctaagc gcagtattctg cctgtgctgc cctggtgtgc atcaactttcc 360
actctggttc acacctcttg ggctttcctgg gcgttgtgct gccttgcagta tattctccatg 420
cygtctgcct gcacactatc cagaacccct aatgtcagcga gcgttgtgct ggagctgctac 480
ttgccacact gccacatcct caagggcaca acctcagtt cccagcaccc cacaacaaaaa 540
tccctagcgc ccattattcc agcctgcggg cctgcgtggta taccattcct ttcacaggctg 600
ggcctgctgc acagggcttg tggcggcagc caccggcggg gcgtgtgttg cacaaccatt 660
gtggacacac ccacaagcta gcgtgtcata cccgtaaacc cctccatgtt ccctgggttg 720
cccctggtgg cctattcatt cttaacaccgc gtgattggca acacacacct agtcatgtgt 780
cacggccag ccagggcagcc ccagagagtg acagtggtgc acatctggca cacaacaggg 840
agyccctaca acatcacctt cgcagggcgg cttgctccag ccctgccccg cggagctccc 900
gtcttcagtg ctcgttgtcc tgtcctcttg gcgtgtggcc tgccttacca cgcctgacgc 960
cgtgtgcgtc gccttgcctt ggtgaacag tgggtctcgt tcctttcctct tgtctttcacc 1020	ttttatatc cagctttccgc atcgccgaagc ctgctattcc ctcctcttcc 1080
tcataacagt gttctgctca cttcgccaaa cttctggccc atgtcttctgc ccagcgttgcc 1140
cctgggtgcc gcaccaagag cccagagagc cccagacgtcc ggtgcagcagc 1200
tccagcaacc atgccttcttc cagccagcgc accagggaga ccctgacagc ggcgcgaagc 1260
gctaaactct gcatcctgac ggcgcagttt tttgcacagg atgtctttcc cgcgcagggg 1320
gcgctcgcct cgtggattgg gcgcagaggg gtagctgtcct gcgaagaagcgt gcgcggcagt 1380
ctggatgaaa tcctagcga atatcagggc aaactgacg ttcgaaaact gaaacatcgat 1440
caaaaccttg gcactgccc gaaatacgcc atctcgggtta tcctcagctct gctgctcttc 1500
aaaaacggtg aagttggccgg aaccaagggtg ggtgctacct ctaaaggtca gttgaaagag 1560
tcctcggac gtcacgcgag aacggggcgca gtattaaag atgacgatga caaataataa 1620
gtactc 1626

<210> SEQ ID NO 171
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 171

gtitaagtcga tcctagctat atctcg 26

<210> SEQ ID NO 172
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 172
cagcagcggat aaccaagccc gaaactagttg ggtactctcg g 31

<210> SEQ ID NO 173
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 173
cggggtcgtcga atagaaaaaa aacaggtgca cgcactctcg cgc 43

<210> SEQ ID NO 174
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 174
gctgtgctc ggctcctcgc atacgcaacg cacaagaaaga aaggggtcag ttcg 66

<210> SEQ ID NO 175
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 175
cagggtagct cgggcaagca gggc 24

<210> SEQ ID NO 176
<211> LENGTH: 26
-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 176

gtacgagggtc tccoggtgcgt cgctgg 26

<210> SEQ ID NO 177
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 177
cgcgcatacg aagcgcaccc tcggaatcgc aacggragg gcc 43

<210> SEQ ID NO 178
<211> LENGTH: 73
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 178
ggggoggtac cttgtcata gctcatttta taatctcggc cgcgcacagg ggtctcccgg 60
gtggcgcgtg tgt 73

<210> SEQ ID NO 179
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 179
ggggoggtac ctttatattt gctgctgca ttttttaat cttgogggcgc g 51

<210> SEQ ID NO 180
<211> LENGTH: 79
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 180
cgcgactaaac gacagtagt gtttctgctt cgggtcctgc csaatcatc gaagggcogca 60
cctcgggaatcg cagacaogggc 79

<210> SEQ ID NO 181
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 181
cgcggtcga catgaaatc aaaaacaggtg cacgcacctc gcatatatc gcattaacga 60
cgatgatgttt ctcgctctgc gc 82
<210> SEQ ID NO 182
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 182
ccgccgacga taaaatttt cacotgacg acg

<210> SEQ ID NO 183
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 183
gccgccgaag ttcaagctca ggaactcttt caactgacc

<210> SEQ ID NO 184
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 184
gcggccgcaa ggcgataaaat tattcaactg actgacg

<210> SEQ ID NO 185
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 185
ggctgtccgg cgcgtccatc atgtctttta taatgcc

<210> SEQ ID NO 186
<400> SEQUENCE: 186

<210> SEQ ID NO 187
<400> SEQUENCE: 187

<210> SEQ ID NO 188
<211> LENGTH: 2465
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 188
gtgcgacattg ggcaaccggg gaagggcagc gctttcttgtc tggcacccaa tgtgaagcat
ccgccggac acgactccac gcagcagaaa ggaggtgtgt ggggttgtgg gatcgccac

<400> SEQUENCE: 189
tgtcatcttc ctactgtcct ggccatctgt tttggcaatg tgcgtgcatc cacagccat

<400> SEQUENCE: 190
<210> SEQ ID NO 189
<400> SEQUENCE: 189

000

<210> SEQ ID NO 190
<211> LENGTH: 2485
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 190

gtcgacatgg ggcaacgagg gaaogcgagc gcttctttgc tggcaacccaa tgggaagcat 60
gcgcggccac acagcgctac gcagcgaagg gcagcgggttg gggttgaggg cgattggcattc 120
gtgatgtgtct tctcgctctt ggcagccttg tttggcataat gcttgtccat caacagcatt 180
gcgatgtgtcg aaggtcagtac gacgctttca cagtacctca ctagctctct gcgcctgtgtc 240
gatgtgtcac ggactgtgccc atggggtcgcct tttggttgcgc ccctattcttc tattatgatgat 300
tggactccttt gcaacctcttg gtcgaggttt tggactcctca ttgatgtgct gttgtgccac 360
gcagccttgg agacgctcttg gtcgatgca gtggtatgct actcttgccat accctctcctg 420
ttcgatgcc agacgcctcg gacccggagat agccagggcg tcacccgctct 480
tttgtgtcag gcctctctctt tctttgcccc accgactgac accgtgtaag ggccaaacac 540
caggagcaac tcaactctga tgccaatgac gcctgtgctgc accctctcctg gacccagcc 600
tatgccatcc cttccttccat cgtgctcttc tgcgttcccc cgcgtgatcat cgtttttgtc 660
tactccaggg cttccagatg gcggaaatgg ccgtggcaga gattgaacaa atccgagggcg 720
gctccatgc caggacacat tagccaggtg ggacagggcg gcccagggc gtagctctgc 780
cgcagatctt ccagcttcct ctagggaggg ccataagggc tcaagaggtg aggcatcactc 840
atggggccct tcacccctctg gtcgctgccct ttctctctct gtctactctt gtcgctgtac 900
caggagatac tcaagctggc ggaagtttacct gcttcttcat accgttgatg tctccgctat 960
tctctggct ctccctcacat cactagccggt gacccagatt tccgctctcc ccctgcggatg 1020
tctctgtgcttgcc gagtcttgag cccgatcgct aaggtctctgc cggagaagcg 1080
aaccagccgg gccgactgcgtg atatctcagtg gacccagagag aagagatctg 1140
gagacgccttc ccggagcgct aagttccttg gcgctatagag gctctgtgcc tccgataact 1200
attgattcct acaggagggct tttgtagcata aacagctcagt ctctgtacgct gcggacaagcg 1260
gtgcaacacc tgcgtaaata atcaagggg cccctcagat ggctggtcct cggagaagtta 1320
agccagagcc cccggcagtt cggagggagg cgcctgtcga gcggcaacaa aacagctgag 1380
agcgacgctca gcggacaccag caggtctccgg gcggcacaagg cggctgtgctcg tctctgggt 1440
tctgagaatc tggtaaagac accattgtgg aacagagtggag ctctgtgcat gtttaatggg 1500
ntttgagaca cggttgagag gcggacccag tggcggctcg aaaaacaacg itgaaaggg 1560
cgattgaaac cctcttggtg ggcagctcag caacggccgct tggagagc tggcaacacc 1620
cgcgacagca gttgcagctag gactactccc tgaaggtgat gacctttgact 1680
-continued

tcctcccgag attctatgag cacgcaaaag ttcttgggga gtagaagaga gtagtgcct 1740
gtaaggaac ccgcaaaag tttcagtgct tttcagttgc ccgcatcttc ccggaacaag 1800
tcgtgagtt gcatagtgtgc cgagcatcag gacacgtcct tcctttggtgg 1860
ttcggctgg cttccatcctt ccagggagcc cctgcaaaac tgcgtatcttgt 1920
tgagatgctt gtcgcaagcg cacagactgc gatgaggttg cccagcctcct 1980
tggttagcc tgacagctgt ggtacgctc ctatgttgctg ctttcgcctc gccgtgctaagg 2040
agcagagcc cgcgctagag gcgttaacct gcctcgtgcctg cttcggaggcc 2100
tgcggacacat cccacagcct tcgtccgctac ctcctggtcttc tgcgacattc 2160
ttcgctggga acaagagtct tccgcagact ctcgagctgc actactctctg 2220
agctgtgctc cccgcancc gcacgcaggt gtagacgctcac cgccgagaag aacctcgctg 2280
gagagagtct cgagctgtgc atgagagctc tgcctcgcag tgcaggagatgaggtccgctc 2340
atccctgctg gctgagctgg tgcgctacg cctcagcagctg ccgccgattg ggggggattg 2400
tctccgata ctcgctctg cctcagcata ctcgctgct cccgctggtcctt ccccgcctgc 2460
"<210> SEQ ID NO: 191"
"<400> SEQUENCE: 191"
000

"<210> SEQ ID NO: 192"
"<211> LENGTH: 1152"
"<212> TYPE: DNA"
"<213> ORGANISM: Homo sapiens"
"<400> SEQUENCE: 192"
ctcgagatgg gtgctctcgg gaacagttaag acggagacc agggcaacga ggagaaggcc 60
cagctgtcgg ccaacaaaaa gatgggagaag cagctccaga aacgccagca ggtctagcctg 120
gccagccacc gcggctttgt gcggcttgc ggtggtacct gttacgagctc gttgccgagt 180
cagatgattc ttccgctttg gggcagaggt gctgcttgcttgtactgtcg tcatctctctg 240
cagccagcc gcacagccac gggagaggct ggagagagct gcttgccgct tggcgaggcc 300
tctctgccc cccgtgggagtt gcacacccag gagaaccagt ccaggttggactt ccctctcctg 360
agctgatgac agctgtcctgc accctgtggt cctctctgtg gatttgccat cttctctgtg 420
tctctggtggag ataactggagt gcgctgtgtg tggctgctgc tctggtggtc tcatctctctg 480
gactgtgcc ccacactgc cggcaagcttg gcgctgtgtg gacagcagct gatggtcatc 540
agctgatgac aatctgtcctgc gctgctttt catttttttg ggcttggcgtt cccgctcagg 600
cagctggtggag aagtcagccag aggctttggt ggcctgcgtg ggcctgtgctg cggcgctgct 660
aagctgtggtggag ccattgactc cagctgtggt ggtggtactc ttctcttctg gatggtcatc 720	tcaacattgt gatgtgttgg gaacactgctt accgacgcct tccgtgctgtg tgcgctgctg 780
tcggagtctg gatgtgttgg gaacactgctt accgacgcct tccgtgctgtg tgcgctgctg 840
taagctgtggtt gatgtgttgg gaacactgctt accgacgcct tccgtgctgtg tgcgctgctg 900
tccaagtgtg gatgtgttgg gaacactgctt accgacgcct tccgtgctgtg tgcgctgctg 960
taagctgtggtt gatgtgttgg gaacactgctt accgacgcct tccgtgctgtg tgcgctgctg 1020
-continued

```
ggagatgggc gtcaactctg ctacccctac ttcacotcgcg ctgtggacac cgagaactc 1080
cgctcttggt tcaagactctc cgctggacac attcagcgcga tcgcacctcc gcagtcagag 1140
cgtcctagat 1152

<210> SEQ ID NO 193
<211> LENGTH: 1194
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 193

tctggagatt ggtcgcctggt ggacagtaag accggagacc aggcaacgca ggagaagccg  60
cagcgtgagg ccaacaaaaa gatgaggaac agcgctgacg aggacaagca ggtotacgag 120
gccagcacc ggcgtggttt gccggtggt ggagatgcgg gttaaagcac caagtcggaag 180
cagatgaggg ttcgcccttt caaatagttt aatggaagag cggcggtgagg ggaccgagcc 240
gctgcaaggg gcacacggga tgtggagaag gcacccacac tgcaaggacat caaacaacac 300
cgctgaaag ggaggtgacgc catgtaacgg gccagtcgca acctgtgacct cccggtggag 360
cgctgcaacc cgcaagacc ggctcgagtg gcactacatc tgagtcggtg gaacggtgct 420
gactttgact tccctcccaag atcttcttgc gcagccaaag ctctgggtag ggatgaagg 480
gtgctggtct gctagaagag ctcaacagcg ttcagagttc tgagctgtgg ccagctgttc 540
gcttgagcaag tcgcgaggtc cagcagcggt gcctatgtgc cgacgtgacat gcacggtcct 600
cgctgggctg tcctctttct ggaagaggt ccagcggagc gtaagctgtgc caaagtcgg 660
ttcctcctgt tgcgcgtgag tgcggagcg gcgaacgacc gcagctggac ccaagttgacc 720
aacagtga cttgctactt ctgggctggt ccacagcaga gccacacagc gcttagcagg 780
gagggacacc agacacagcgc ccggcgctag gtttggacac tccttcaagag catctggaac 840
aacagatggt gcgcggagct ttcctgttac cttctccctc cacaagcaga ttcgctgtgg 900
gaggaagtc tgcggctggaa atcgagattt gagcactact tcctcaaaatt gtcgctgtac 960
acactctctctt ggtgggtac ctgctggatc ttgctctctc aacaagcaga ttcgctgtgc 1020
tgacactctcttgagctgacgac ccaagacgagc acagctgtct ccgtgcctgtc 1080
tggacatgct aggactgctc ccggtggagc gagagacagcg cagcagcagc ccggggccag 1140
ttgccgcag ctacacatcc gcagctacgc gcggagcagg agagctgcgg 1194
```

<210> SEQ ID NO 194
<211> LENGTH: 1089
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 194

tctggagatt gcctgsgact catcacgtgcg tgcgctcttga gcggagacgc caagagaacc  60
cgcggcagat acgcgacagt cgagccgacag ctccgacggcg acaagcgagc cgcggccgg 120
ggcgcgtgac gtcgctggtg ccggcgagct ggaagctgca caagcgcgct tatcagacgc 180
agtggatcact tcggccctctt gatagactct gatagacatg caaaggggc cccacagctg 240
gttactcag atcacttcaag gccctgctgag gcacacagcag gcacacatg ccacactcaag 300
atccatacg aggatcagcgc cattcagcct ccagccacaag tgcgctcttg agttgcagtg 360
gagaaaggg gtctgctttga gaaatctgat gtagatgac ccagacagtt gtagagatg 420
-continued

cctggaaatcc aggaatgcta tgatagacga cgagaatact aataatcctga cttctacaaaa 480
tactatotta atgcctcggga ccggtgagct gcggcctcct acctgcctac gcacaagat 540
gttggttagc tcgaagtcgt ccacoaccag ggatctgaat atcccaatt gcccaaaatgtttcaaatagtt 600
gtcatatttta cacactggcag ttgaggggcc cagaggtgca agaagagaaga atggatccac 660
tgartttgaa aatgctcggc ctagggtgaa ttgtaatggt ttaatgtaaa agatctaggtg 720
cctgcgagct cagacaatgca gagacgaaag caggtctttt tagaacaacct 780
ttatactac ccctggttca cgaactctcg gttatcttct ttctaaacaa aaaagaatgcttt 840
cctagaggaga aatactgtag ttccacactaa tgcgtactt ctcagagaata tgcgagaa ccccttg 900
cagagagttt cccccagagg cccgaaatatc atctctagaa ttccgtgaa cctgaacccaa 960
gacagtgaca aataatcact cttccaccttc attgtgacca cagtgcccag gaaatctgagc 1020
tttgtcttttg tgcgtctga ccaaccac gcctacgtgca acctagagga gtaatactctg 1080
gttctgat 1089

<210> SEQ ID NO 195
<211> LENGTH: 1077
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 195

cctgcagagtc gcgagccgct gacgcggcct gacaggtgag gcagaacactg cctctataagtt 60
atgccacaag aacgtggggg gagccggagag aggccgggcag ggagggtgaa gttgtctttg 120
tggaggcttgc ggtaggagcg caagagcacc atcgtcgcag agatgagatg catcaccag 180
gatgcgccact cggagtcgag atggccgagg tacggggtcc tgtctcagc caacaccatoct 240
cagcctctaca tgcagatttt ccacaagctgt gacagctctg gaaaccgcag tgcgagacc 300
tccagagcc gcagagccac gcgcggatct ggcgcctgct gcggagcctgt tcgggagcag 360
gttgctcctgg atgactgtcg gcggagcctgt ggcgggaccc gcgtgctagct 420
gcggctgggg gcgcggagcgg ccattgtagc cagtgctctg gcggtcagcc gcggagcctgt 480
gactggaggc gttcagccag atcggagact aaccccgac acgcgcaggt cgtaagggcg 540
cggcgagagc cggtggagat cgcgtggcgg gcaagagctt ccactgtgg gttggtggttg 600
atgttagtg tggggtgcgg gcaagagagct ggtgctcaga ccccttaggt gttgttggt 660
gtccagccca ccttttctcg ccgtaggtgt ttgctgcctg gctggagagc gatctgctgtg 720
ggcgggagc ggccggagct ccagaggtgg cattctgtct aaagtcctagt gttgtgctgtg 780
ttgctcagc acaacgcttc ctattcttct ccacaagag accagtttta cagggagagcc 840
atccacacca gtcctcgtgac ccgtgggttt ccgctgcttc ccgggctgatt ctttagtcgaa 900
gaggcacgca gcctaatcct gagtggttaa cccggggaca aacaagcagc 960
gagatctaca cggtttctcc gcgggctttg cgggcctgtg cctgggggcag 1020
gcgccgcgg agtcgcctcg caagacggcct ccgctgcctt gatccg 1077

<210> SEQ ID NO 196
<211> LENGTH: 1155
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 196
-continued

cctgagatgt cgggaggtgt ggggaccect aagcgctgcc tgcgcgcegc cggagccggc 60
gggggccgc gcgcggaggc gggccggcgc gcgcgcagcg cggagccgca ggcgcgggag 120
egtgagcccg acatccagcg cgtgctgccc cggagccgge cgcggcctgg cgcgcctcg 180
aggcatcctgc tgcgggaggg ggcgcggcgc gccgaaggcc cgtgccctccac gcagacgcgc 240
atcatcacc gcgcgaggtt gcacagagaa ggcgtgtctgg agttcggcga cccatccctc 300
gacacacacc tcaagggctct aagggctcct cggagtgccc gagataagct tcggattccct 360
tggccgttta cgggaaatgc tggaggctgga atggcttctg tggccatttt gaaacaggcc 420
ggcggctctgc tggagccggc cactctccac cttgtaagct ccggcccctgg gcgtacccttg 480
agggattctgc gacatcaggg aagtttcggc aagagaggag agttcctagct ggggagctcg 540
gtgaagctact gtctgagcctt cggagccggg atcgcccgac tgaattacctt tgcagtaag 600
cacagatctc ttcctgcccag gcagggccgc atggccagcc tgcagatttt cctggattttt 660
aagacagtcc accattatgt ggtggagttgc gcggcccgcc ggttcccgcc cggagttgcg 720	tccagctgct tggagggatt cagttccccc cacgttcctag tggcccagtg ccagttcggc 780
cagtctctcttg gggagacac gcgggtaccag cgcgctgctgt agttctgaga caaagcctcgag 840
accactggca aacacaaactc cttcttccac gtctttctttc tcacgtctcag cggagccggc 900
gccttcctgttg gggagaggtg cagacccctg agataacaaaa aagttgacctgg cggagctggc 960
gggagccacc ctcagctggtt ggggagcttg ccgcagctgt tagtggctgtg ccagggaggtc 1020
agagggccgc gacgcaagcc aacttccag ccacacaac cccgagctgc cagcgagag 1080
gtctcctgtcg tgcagctctgc tggaaagacc aacagttcgc agggcggcct gggagccgc 1140
atgcgtcgaga tggat 1155

<210> SEQ ID NO 197
<400> SEQUENCE: 197

<210> SEQ ID NO 198
<400> SEQUENCE: 198

<210> SEQ ID NO 199
<400> SEQUENCE: 199

<210> SEQ ID NO 200
<400> SEQUENCE: 200

<210> SEQ ID NO 201
<400> SEQUENCE: 201
<210> SEQ ID NO 202
<400> SEQUENCE: 202

000

<210> SEQ ID NO 203
<400> SEQUENCE: 203

000

<210> SEQ ID NO 204
<400> SEQUENCE: 204

000

<210> SEQ ID NO 205
<211> LENGTH: 3307
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 205

gtgcacagg ggcacaccgg gaacggcgcg gctctttggc tgtgcacccaa tggagcagtc 60
ggccgcacc acagcgtac gcaccaaaaag gccagggttg ggttgtggtgg cacggggaccc 120
gtcagtgcttc tctctggcgc ggccatcgtg tttgcaaatg ttcgtgtcatt cacacgcatt 180
gccagaaggg agccggctcag gcgccacacc aactctcaca tctctctacatt ggcctgtgct 240
gatctggctca tgtggctagc aagttctgccc tgtggggcag cccataattc tgtgaaaaatg 300
tgagatcttc gcacatccttc gtggagtgttt tgtgcttcata tgtgctgtcagc 360
gccagaattg agacccctttg cgtgatacgc gttgagagtc attttgccat taccctcactct 420
tcaagatcct aagagctagctg gacaccaaat aagcccccgg tgtcactttct gatggtgtgg 480
atttgctgcag gcctatcctt cttcttcctcc attccacatt gactttcagc gagccaccac 540
cagggagcca tcaagctgcta cccctgagag acctggcttg taccctcccc gcaccacacc 600

tatgcaatttg ctcctttcct cgtgcttcttc ctagttcccc cccttgtgacgc gttttcgtcg 660
tactcaggg ttttctaggg gcggggaaag cacgctcagca gattggaacc atctgagggcg 720
cgctttccct tgtcagaaatc gacgcaagt gggcagggcag ggctgctgctc gactggctgc 780
cgtccaccttt ccaagatgctg tctggaggag caaaaccctt caaagctgcc aggcatcacc 840

tagggacactt tattaatcttg cttgctgccc tttcttctagc tgtaccatttt gctgctgatac 900
cagacaataa tctgctcatt gcagttctac atctccatcc atcagatagc ctgggtccatcc 960
ttctggttctct tgtatctcatc gactctgccc agccaggatt ctggagtttc cttctgttcg 1020
tttctgtgc tgtcgtgaga tttcttggaag gctattgaag ggtgtgtgca atgctgtcato cagcaaggcc 1080
aaccagaggg agccagatgg atactcagtg gacacaggag aagaaatata actgtctgtgc 1140
gacgacagcagagcctgctt gcctgtggca acgtgtaccc tgggtactac gatgctcatcct 1200
attgtagcttc aagggagagg tgtgtgatcac aactcctcag tagctagacg tcggcagaccg 1260
gtccaaactgc gctcgggac caaccacaaeg cagatcagca tgtgtctcat tgtgactaaac 1320

ttctagttg cttgaaaaat ttcctcctgct ccctcaagctc atactctgtgct gacaaaggg 1380
gatagtgaag agatcattcg attaggcagc aacgaaagcc gaattctttg gcgtgcggc 1440
cacgtcaca acgaggctat ttctgcataa ctttggcatag actttgttgy gcggagcaaa 1500
gttttgtga tagatgatcc cagottaacc aacgacattt cgaactctgcg caaatagtcc 1560
aacaatgca caaatgtccc aacataagtc aaaaaggttc cgaacgcgag ttaccaattg 1620
atcgccccag tggaaacggt tgaagaagag atggctcgcg aaaaaggaag tgtcctagac 1680
atctcttaat atacaaagag gcccctgaga ttgtgtcctt cgggagaagtt aagacccgagg 1740
accaccgcaaa egaggagaga gggcagctggt agggccacaa aagatcgcag aagcagctgc 1800
agaaggacaa gcacgctcag ccggccacgc accgctctgc gctctgtggt gctggagaa 1860
tggtaaag acacagatgta gatcctgca tgtataaggg ttataatgag 1920
aagcgagga acgcaacaa aacgacagca tcaaaaaacc otggaaaag gcattggaaa 1980
ccatgtggc cgccattgcag aectgtgtgc ccctcctgga gctggccacac ccggaacc 2040
agttcagagt ggttccctac ctgtagctgg gtaaagcgtgc tcattctgcc tcctccctcg 2100
aatcttactg gcctgcaacag gctctgtgag agatcggagc ggtgcgtgcgg tgtctagga 2160
gecctcaacg tttacagctg attacagtgc ccagtaagtc cctggcaacag atcgactgta 2220
tcaagccggc tgacactatg ccagagcgata agacacgctg tcgctgtcgc gcctgtactt 2280
tcggatcttt tggacacaag ttccagcttg aacaagctaa ctccccacatg tttgacgtgg 2340
gtgcagccag cgtaaagtgca cccctagttg cccagatgcttc atgcgcatac 2400
ttctggtgt gcgcctagcc agttacaaac ccctgtcctcg gcggacaaac cagaacacc 2460
gctgcggag gcgtctgcac gctctccaga gcctcttgaa ccacagatgg tgtgcgcacc 2520
tctctgtttg ccctgctccc ataccaagga atctgtgcgc tggaaaaagt cttggtaa 2580
acattcagat tggaggactac ttttcgcataa ttgtgcctga caacttcctc gcgggtggaa 2640
tccctgagcc egagagagcc ccggcctgtg agcgggcgcac tatacttcatt cgcagatgat 2700
ttctggtat cagcaagctgg gccggcagtt ggtgcgtccta gctgtactct ctttccactt 2760
gcctgtgga cactccagaa atccggccgt tttcgcctga cccgctcgtc atcaagcggc 2820
gctgtcact cccttactag gcagtcctag tgggtggttgc atgttcagaa ccgttcacaa 2880
tgatcctat gtttactaaa tttacttttg otggaaaatt tattctgcac cctctgaagct 2940
atctcttgat tgacaaagag gatagtgagc agatcctcg attaggcagc aacgaaagcc 3000
gaattcttttg gcgtgcggc cacaagctcata cctcggctaat gatttgcatt 3060
acattttttg gcggagcaca gtttttggaag tcggtagctc cagottaacc aacgacattt 3120
cagactctgcg caaaaagctgca aagattcctg ccacaaacctgccc acatatacgc aaaaaggttc 3180
cgaagcgccgg ttaccaattg atgccccagc tggaagaagag atggctcgcg aaaaaggaag 3240
aaaaaggaag tgtcctagac atctcttaat atacaaagag gcccctgaga ttgtgtcctt 3300
aagggg 3307
SEQUENCE: 207

000

SEQ ID NO: 208
LENGTH: 3284
ORGANISM: Artificial Sequence
TYPE: DNA
FEATURE:
OTHER INFORMATION: Synthetic nucleic acid sequence

SEQUENCE: 208

tctagaggt cggagctgaa gttgaaacgg gtttaccgtat aaaaacgaa aatgataaa
aaggactaaa tagtatattt tgatgttgta tttggtattt caatataaac aatctatttt
acattatccttgctgtttaac ctaattctatg atttatttaacc aagggagca tcttatatg
aacacattga ttacggtacctcagcggtgtta ttgacgagct gccgtaactgc ggaagacccct
aggaggggccc ccgtctcctactacgccg catcgggaga cggcccccctg ttcggaagctg
agctgttactt tcgtggcaagc agaaagctgt cggaggggtg cggcttctccctctctccct
 gagagtagatt cgggtgctctgcttacccgc cgtggtctgctgctgcctgg cgtggtgctcc
 taccacacactggcctctactgcctggtccccgaggt gaacattgc ttcctgcttttacatg
tcgccctgcctgttttacttctgt ccacgtggctt accggcgtggctgctgt ttctggttgg
 cggcttccgctg ttcggtatgt cggcgcgcctg cgggttctgtg tttggtggttcttgcgttgc
tcgggttctg ttcggtatgt gcgttcttctctgctg ttcggtatgt gcgttcttctctgctg
tcgggttctg ttcggtatgt gcgttcttctctgctg ttcggtatgt gcgttcttctctgctg
tcgggttctg ttcggtatgt gcgttcttctctgctg ttcggtatgt gcgttcttctctgctg
cctctggttact agggcgggct gttctgggac tgggttggtac agtgcgtgat taaatatgat 1920
gaaaaagcgc accgggtggtc ggtctgggac ggtgtttttg gcgtatagcc gcgcgtgcgc 1980
cagttcgtcta tgaacggtct gttttcgtcc gcagcgcggc cgagcgccgc gtgaagggaa 2040
gcnnmaacc agcagccggg tttcccgttc cgtttttcccg ggcmaaccac ctgaagtcgc 2100
agcgatgcc tggccgctca tggctctgag tgggtctggtt gccggttgggt ggcggttgat 2160
gttacggtgc tggcagccgg tgaagcgcct cttgatctgg ctccacaaag taaaactcattg 2220
atatggaatc ctgaactacg gcacgcgggag aagcgcgggg aatctctgtct cacagtaacc 2280
gtattgacac gcacggcgcgc cgacggtgct gatagctgcct ctcggtgctac gctagccgc 2340
ttgctgtgg ccgaaaaacc cagtgtgtcg ctctcgccgct gctccacagc cagccctcag 2400
tctaccacca gcggaaaagt tttttgcttc gacggtggta ataaggtttg gccaaatattaac 2460
cgcgtcagcg gttttctcct acagatgtggg atggcgcata aaaaacacact gctggacgcgc 2520
cctgccagac agtcaccccg tgtgacgctg taaacgcaca ttagggtaag tcggcgagcc 2580
cgattgacac cttaagcgctg ggtgacagcg tgaagcgcgg ccggcgctta ccaggccagaa 2640
gcagcgttgt tgcagcgtgc gcgcgtatact cttggtctag gctgtgctgat tggcgtacgc 2700
ccagcgtggc agctcaaggg ggaaaaactta ttatcagccc gggaaaaacta cggattggtg 2760
gtgagtgttgc aatagcggct taccgggttg gtgtaggtgg gcgcgtccac gcacgcgtcgc 2820
ggggcggatt cgcctgctcgg cagctgctggc cagtcggtct gcgggttgaac gttgcgtcga 2880
ctttggcggc aagaaaaacta tcggggctgt cttagctggc cctgttttttg gctggtcgatg 2940
cctgcatttgc cagagtctgtg taccggctgct gttttccgga gcgcggagcc ctgctggctg 3000
gggcggtgc ggtgagtgcata tggcccaacac cagtgccggt ggcagcttctca gttggctac 3060
gggtcgacta tggccacacgc actgtgggaa acacggtcag gccatctgct gcacggtcgaa 3120
gaaaaagcgc tgccagcgat gcaggggttc ctagttgggga ttgctgggca gcacgctcgg 3180
agccggtcag tatttcggag attccagctg aggcgcgggc gcacggtcag cccgctggtc 3240
ttgctgtaaa aataatacgg ccttttacgc aagagcgctg tga 3284

<210> SEQ ID NO 209
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 209

ggggcgacc gcgggggcg ggcgc 25

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

<400> SEQUENCE: 210

gcagtgagtc atttgcatca cattctcc 30

<210> SEQ ID NO 211
<211> LENGTH: 37

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

<400> SEQUENCE: 212

tgctgtagtc tttcctctgc gcagcggcgc 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 211

cgctggctgc atggggsacc ccggsgasgg ccgcggcc

<210> SEQ ID NO 212
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 212

ggtctgacgt gcaggcttgt gagcgtcttg ccagcgtcata gcagtgagtc atttgtacta caattcc

<210> SEQ ID NO 213
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 213

gggctggcttc gggacagtac agacgagg

<210> SEQ ID NO 214
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 214

gagcagctcg tactgacgaa ggtgcagtcg

<210> SEQ ID NO 215
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 215

gagggccttc gagattggtc gtctgggaa cagtaagacc gagg

<210> SEQ ID NO 216
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 216

cctcttagct attactgtag agcagctctg actgacgaag gtcgatcg

<210> SEQ ID NO 217
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 217
<br> ccatgatga gcagcgtga tgcagaagg tgcagc 37
<br>
<br> <210> SEQ ID NO 218
<br> <211> LENGTH: 26
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 218
<br> cogggtgtgt goggacocct acgcgc 26
<br>
<br> <210> SEQ ID NO 219
<br> <211> LENGTH: 28
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 219
<br> ctgcagaagt atgtcctca ggttctcc 28
<br>
<br> <210> SEQ ID NO 220
<br> <211> LENGTH: 41
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 220
<br> gcggacctca gatgtcggag gtgtgcgga cccctcagcgc c 41
<br>
<br> <210> SEQ ID NO 221
<br> <211> LENGTH: 39
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 221
<br> gccgcatgca tcgtcagcat gatgtcctc aggttctcc 39
<br>
<br> <210> SEQ ID NO 222
<br> <211> LENGTH: 28
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 222
<br> gactgtgcg tcatacatgg cgtgtg 28
<br>
<br> <210> SEQ ID NO 223
<br> <211> LENGTH: 29
<br> <212> TYPE: DNA
<br> <213> ORGANISM: Artificial Sequence
<br> <220> FEATURES:
<br> <223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<br> <400> SEQUENCE: 223
ccagattgta ctcttcagg ttcaactgg

<210> SEQ ID NO 224
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 224

atgactctgg agtcctcagt ggctgtgctgccc

<210> SEQ ID NO 225
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 225

ggcgcataca tgacagatagc tcaccttc acggtcaact ggccc

<210> SEQ ID NO 226
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 226

ggctgcacc gcgtgccgag ggcacagg

<210> SEQ ID NO 227
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 227

ctccaggtt gtctctgtag agtcacatggg

<210> SEQ ID NO 228
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 228

aggggctgca cagtgagatg cggacaggacg

<210> SEQ ID NO 229
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 229

ggcgcataca tgacagagcc gcagtccttc aggtgtctt gatgatgac atcgg
<210> SEQ ID NO: 230
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 230

gggctgcctc gggsaacagta agacggg

<210> SEQ ID NO: 231
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 231

gagcacgctg tactgacgaa gggtcagtc

<210> SEQ ID NO: 232
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 232

atggggctgcc tcgggsaacag taagacggag g

<210> SEQ ID NO: 233
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 233

gggccatcga tgagcacgctc gtactgacgga agtgctgatc

<210> SEQ ID NO: 234
<211> LENGTH: 107
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 234

gggtgaggg ccctcgtat tattactgca gggctctctt gatttactgc caggtgttgt gacccgcttg ccagctctca gcagtgatcg atttgtaact caattcc

<210> SEQ ID NO: 235
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 235

cctgcaggt tgtgacggct cttgccagct ccctagcagtcagcttgc gagcttcatt actacaattc

c
-continued

.Feature:

Other Information: Synthetic oligonucleotide primer sequence

Sequence: 242

cgggtctgg aggttgtgg tagaagtga aacgggttta cgg 43

Seq ID No 243
Length: 60
Type: DNA
Organism: Artificial Sequence
FEATURE:
Other Information: Synthetic oligonucleotide primer sequence

Sequence: 243

cgagggcag tgaatcgtat ctcacctata atgctccott tgccttaacag 60

Seq ID No 244
Length: 27
Type: DNA
Organism: Artificial Sequence
FEATURE:
Other Information: Synthetic oligonucleotide primer sequence

Sequence: 244

cgattacc gcattcactg gccgtcg 27

Seq ID No 245
Length: 27
Type: DNA
Organism: Artificial Sequence
FEATURE:
Other Information: Synthetic oligonucleotide primer sequence

Sequence: 245

cagaccacac tgtaattggt agcgacc 27

Seq ID No 246
Length: 35
Type: DNA
Organism: Artificial Sequence
FEATURE:
Other Information: Synthetic oligonucleotide primer sequence

Sequence: 246

gttaaagcac atgattaagc atctactggc cgtcg 35

Seq ID No 247
Length: 77
Type: DNA
Organism: Artificial Sequence
FEATURE:
Other Information: Synthetic oligonucleotide primer sequence

Sequence: 247

ggcctcttag aataagcct tttggttag ggcgtttatt atttttgaca cagaccacac 60
tgaatggt agcgacc 77

Seq ID No 248
Sequence: 248

000
<210> SEQ ID NO 249
<400> SEQUENCE: 249

000

<210> SEQ ID NO 250
<400> SEQUENCE: 250

000

<210> SEQ ID NO 251
<400> SEQUENCE: 251

000

<210> SEQ ID NO 252
<400> SEQUENCE: 252

000

<210> SEQ ID NO 253
<211> LENGTH: 89
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 253
cgctgacgcag cgtcgacgcag cgtcgacgcag tgaattcgc gagagccag ggyygygyy 60
cgatgatgat gctccgctct gcgtctgcgc 89

<210> SEQ ID NO 254
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 254
cgctgacgcag cgtcgacgcag cgtcgacgcag tgaattcgc gagagccag ggyygygyy 60
tcgtctcgc 69

<210> SEQ ID NO 255
<211> LENGTH: 81
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence
<400> SEQUENCE: 255
cgaagccgcc ctgcagcgcct ctgcagcgcct gcgtctcgc gattataag atgcagatga 60
caacaaaat aagcagcagc c 81

<210> SEQ ID NO 256
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
-continued

<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 256

gcccgctag cttattatctt gtcatcg

<210> SEQ ID NO 257
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 257

gtgcacgcct ctcgcttcatt atcgc

<210> SEQ ID NO 258
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 258

gggttttcc atgtggcggg caataacgttg

<210> SEQ ID NO 259
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 259

cgggagatga cattgaat attaccggtg caagcctctcgc cggatttgc gc

<210> SEQ ID NO 260
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 260

cgggagcttg cagcgagctg tcgatacgc caagcattaca gcgaagcgac tgggcttttt ccatagtggc ggcataacgtg gg

<210> SEQ ID NO 261
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic oligonucleotide primer sequence

<400> SEQUENCE: 261

ggccgctag cttattatcttt gtcataccta tcttataat ctcagatctt ggcgtgacc ggcgcctcagc ggccggcact gatacg

<210> SEQ ID NO 262
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 262
cctgactgac gacagttttg acaacg 26

SEQ ID NO 263
LENGTH: 32
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 263
cctttgaca gtcacaccac tttggtggcc gc 32

SEQ ID NO 264
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 264
cgcgctggtca gcgcctggct gacgcggaat otgaaagcga taataattac cactgactg 60
acagcagttt tgacacgg 78

SEQ ID NO 265
LENGTH: 95
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic oligonucleotide primer sequence

SEQUENCE: 265
ggcgctgtag cttaatatgc gcctgctga tccttgaact cgcgcaggtt ctcttccac 60
tgacctttgg acagctgacc caacgtggtt gcgc 95

SEQ ID NO 266
LENGTH: 216
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic nucleic acid sequence

SEQUENCE: 266
gaatctgacc gctttttgtaa ctggtgctaa tggaatctcag aggttttgc atatgaaaat 60
aacaacaggt gcacgcaatcc tggccattatc cgcattaacg acagatagtgt tttcgcgcct 120
ggcgctgoc acaatcagc gacgcgctcg gcgcgcaatc gtcgacgcgc aatactagaga 180
atataagat gcagcatgaca aataataacg ttagg 216

SEQ ID NO 267
LENGTH: 262
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic nucleic acid sequence

SEQUENCE: 267
cataacggtc tttttggtgc ttagcaggac cccgctcataa gaaaaatcag aacagtgacc 60
-continued

taatctcggc attatccgca ttaacgcga tgtgtttttc cgcctcggct ctcgccaaaa 120
taatgaaac cgcctcggac gcctccgtcg acgcgaacag acagcgaattc agaggttaag 180
atgccaatta ataacgtgaa ggg 202

<210> SEQ ID NO 268
<211> LENGTH: 182
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 268
aggaggttct gcataatgaa ataaaaacag gttcaacgcct ctcgcatta tcggcatta 60
cgacagtcat gtttcgtcgc tcggtctcgt ccaaatcat cgaagccgcgc tcggcaggccct 120
cggtcggc gcgaatctgaa gattataaag atgaagagatg caaataataa gctagaggta 180
c 182

<210> SEQ ID NO 269
<211> LENGTH: 182
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 269
aggaggttct gcataatgaa ataaaaacag gttcaacgcct ctcgcatta tcggcatta 60
cgacagtcat gtttcgtcgc tcggtctcgt ccaaatcat cgaagccgcgc tcggcaggccct 120
cggtcggc gcgaatctgaa gattataaag atgaagagatg caaataataa gctagaggta 180
c 182

<210> SEQ ID NO 270
<211> LENGTH: 1980
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 270
gaatccgc gttttttgta ctggtcgtta tgaatcctcag gaggccctgc atatgaaaaat 60
aaaaacaggt gcacgccacct ctcgattatgt cgacagtcgct gttccgcctc 120
ggtcctcgc aacatcgcag tcggtcagttac gcggtcagaag tcgttttta 180
taagcgcggt gtaacagttg tgaagcagat cagctggaaag aagccgagct 240
tgacgctcc gcgaatacct gggcagcatt gcgtttttaac ggcgccagtg gcgtcggcct 300
tggtcattcg cgggacctgg tggggttttc gtttacgctc gtcggtctggc 360
tgaaatcacc cggcagaagtg cgtccagggc agctgtgtat cgttgtttct cgggttacgt 420
acgttagaac gcacagcgtc atgtggtcgtcg gttggtgctg gatcagcgat gttgcggttc 480
taaaacaggt ctcggtcggcg ccggtcagag tccggcagtt ccggctccgcct 540
agagcgggtt gcagagtcgcc gcgtccagtt gccgctccct cgggctccgcct 600
cgcgggtggt accggggttt ttcggcctgc gtcgtctcttgcctg gtcgtcttctg 660
catagagcgtg atctggttcat gcgagcagct tgtgcgttggc ctggtgctctgccg 720
cctgattaaa aacaaacaca tgaatgcaga cccggattac tccatcgcag aagctgcctt 780
ataaaggg gagaacagga tgacatcga aagccccgct gcatggtcc caatcgcac 840
cagcnaagtgg aattaggttg taacgtaact gcgcacccctt aaggtcctac catcacaacc 900
gtctggtcg gttgctgagcg caggttaata cccgcgctcg cggcacaaga aagtgccaga 960
agagttcttc gaaatcatac tgcgtacgtga tgaaggtcttg gaacgcgtta atataagcaca 1020
acgcgtgcgt ggctagaga gcgaatctttt cgggaagaag ttggccagaat ttcagctat 1080

<210> SEQ ID NO 271
<211> LENGTH: 1196
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 271
ccatacctggt tccttggtgc tagcgagagg cctcgtcatg gaaatagaac ccaggtgcgac 60
gcagctctcg attacgcgga ttaacgacga tgtgtttgtt ccgcgtgcgt ctgcacaaaa 120
tcgaggaag taaacctgta atctgctgatta acgctgatat aacgatataac ggtgcgtgctg 180
aagtggttaa gaaattcgag aagataacgg aataaaggt cccctgtagtag cattcggata 240
aactgagaaga aacattcgca caggtggtgg caaatcgccga ttgctgtcag attactttct 300
gggcacaga cagctttgttg gctgaogtca cactctgcgt cttggctgaa ataccccgag 360
acaaagctcg cccaggacagat cgctatcctg ttcctcggca tcgctgacgt tcaacacggca 420
agctgagtgct ttacgcaaatgc ggtgtgtaag cgtattagct gattttaca aaagacagtcg 480
tgcagacac gcacaaaaaa cggggagaag tccggcgcct ggtaaaaaag ctagaaacag 540
aaggtgaag gcgcgtgctag tttcgaactgc aacaaagctc ccacccctcg cgcgtgagcg 600
cctgtgaccgc gggtattgct tgtcaaatag aacaagcggaa tgcacgacc aaaaacgtgctg 660
gcgtggtgtaa cggctgcgag aagcggggcc tgaacctttcc gttgacgtcg attaaaaaca 720
aacacatgaa ttgcagacac ggtacttccca tctcagacgc tgtctttat taaagccgg 780
cagcgtctgc acaccaagcc ccttggtgcgt gctccactc gcacaacccg caaatgaatt 840
atgggttaac gcgaagctg aacgccatgg gcagctgtcata tccggccgt tggctgagcg 900
tgcagccagtt tataaacgcg gcagctcaga acaaagcctc gcggaaagag ttcctcggaa 960
actatcgtg gcctgttgaa ggcttgtaaa aagcacaacgg ctggctgagcg 1020
tagctgctga gttttagcag ggagagttcg cgaagattc acgatgtgc gcaccacattg 1080
aacaagccca gcctctcttc tgcctatgcag tgcggatcgtg agcggcctcg caggcctcg 1140
tgcaagccaga ttcggagatatg aagatgctcag caataaagct agagaaa 1196
-continued

ggattaacgg cgataaaggg tataacggtc tcgctgaagt cgtaagaga aa ttgcgaaga
180
atacogaat taaaagcacc gttgagcata cgataaact gsaagagaga aa ttcccacag
240
ttgccggaac cggccagag ccctgacatt tttcttgagc acacgacgcc tttggtggtc
300
aocgtcaat cggccgcttg cgtgaataca ccccggacaa agcgttcgac gacaagactg
360
atacgttttaac ctgggagctgc gtaagttaca acgcgtcagag gattgcctac cgacatcgcg
420
ttgaaagcgtct atcgctgatt tataaaaca agaaccgcgcc gaaacccgcaaaaaacggg
480
aagagatcc gcgcgctggat aagaaactga aagcgcagag taagagcgct gcgtagttca
540
acgctcaaga acgcagactc acctggcgcgg ctgctgtgcg tcgacgaggt tattgcctca
600
aagatgaaaa aggcaagact gacattacag agcgtcgggg cggatacgcg ggcgcggaga
660
cgggtctgcac ctctctgtgt gacctgttaa aaaaacaaaa caatgaagct gacacgatt
720
actccatcgc agaagtcgcc tttaataaag ggcacaagcc gatgacacct aacggcgcgt
780
gggcgtctgct caaccacacag aaccgcaagaa gtaattatgct gtaaccgctg tggccgaccc
840
tcaggggtca accccacaag cgttctggtg gcgtgctgag cgcagttcat aacgcgccca
900
gcggccgcaaa aagagctcgg aagacgccca aagacaacat tctctgagt cgtgaaggtc
960
tggagctggc ttaaaagac aacgctggg gttgcgtagc gtctgaagct tacaaggaag
1020
agttgaggaa agatccagcg atggccgcaaca cctggaagaa gcggccagcgc gctttctggt
1080
attgcgctgcg tataagagcc cgcgctgcag cctggtgctga gcggcatact agaaggatata
1140
aagattcag cggacaagta taagcgtagag g
1171

<210> SEQ ID NO: 273
<211> LENGTH: 1171
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 273

agagagggtcct gcatatgaaa ataaaaacag gttcaocgtct cctcgacatta ttcgcatta
60
cgcagagctt gttttccgcc ctcggctctcg ccacaatcga agaaggtaaa cttgtaatct
120
ccagattacg cgataaaggg tataacggtc tcgctgaagt cgtaagaga aa ttgcgaaga
180
atacogaat taaaagcacc gttgagcata cgataaact gsaagagaga aa ttcccacag
240
ttgccggaac cggccagag ccctgacatt tttcttgagc acacgacgcc tttggtggtc
300
aocgtcaat cggccgcttg cgtgaataca ccccggacaa agcgttcgac gacaagactg
360
atacgttttaac ctgggagctgc gtaagttaca acgcgtcagag gattgcctac cgacatcgcg
420
ttgaaagcgtct atcgctgatt tataaaaca agaaccgcgcc gaaacccgcaaaaaacggg
480
aagagatcc gcgcgctggat aagaaactga aagcgcagag taagagcgct gcgtagttca
540
acgctcaaga acgcagactc acctggcgcgg ctgctgtgcg tcgacgaggt tattgcctca
600
aagatgaaaa aggcaagact gacattacag agcgtcgggg cggatacgcg ggcgcggaga
660
cgggtctgcac ctctctgtgt gacctgttaa aaaaacaaaa caatgaagct gacacgatt
720
actccatcgc agaagtcgcc tttaataaag ggcacaagcc gatgacacct aacggcgcgt
780
gggcgtctgct caaccacacag aaccgcaagaa gtaattatgct gtaaccgctg tggccgaccc
840
tcaggggtca accccacaag cgttctggtg gcgtgctgag cgcagttcat aacgcgccca
900
gtcgaacac aagagctggcc aagagttcc tcgaaaacta ttctgctgact gatgaaggtc 960
tgaaaggttt taataacag aacocgctgg gtcgctgagc gcgggaacct taccaggaag 1020
agttggccaa agatccacgt atggcgcgca ccattgggaa aacccagtcc gcggggctgt 1080
agctccctgc ccattaacgg cagcctggga ccggccagtc agagattata 1140
aagatgacga tcgacaatga taagctcagag g 1171

<210> SEQ ID NO 274
<211> LENGTH: 392
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 274
tagcaggaggt cccctggcggc ctcggtcgc acgcagaacta gaacgtctaa aatattcaca 60
tcgtcgtacg aagctttggta tacaaggtaca tccaaacgcgg acggggccgtct cctctcgtat 120
tttggtggcag agttggtggtgg ctcggtgcaaa aagatgcgcgg cagatccgga taaatctgct 180
gacgatatat cccgcaaatg gcacgttaag caagacaacta tcgatcgaaggtc cccctgtc 240
gccgcaaat agggtcagttc ccggtaatgg ccagcttcaagat gcggcattat 300
gccgcaaccc aagttgctgg ctcggtctagtt aaggaacactt gttcgggtga 360
aagatgacga tgaagacata ataatgctagaa g 392

<210> SEQ ID NO 275
<211> LENGTH: 426
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 275
gaatccagcg gttttaagta cctggtgatatt cgaatattcag gagggtcttg cagggctcggat 60
cgcggcgaat tctgaagacg ataacaattat tcaactgact gacgcaagtt ttgacacggaa 120
tgtaacttcc ggcaggcggg gcatctcctgt cgatttctgttg gcagagttgc gcggccttgt 180
caaataat gcccctgattc tggatagatgt cagacgcaaa tatacgccca aatagcaagt 240
tggcaaaatt gcacctgctt cccacccctgg acattggcgc acggctcttg tcaagctctgct 300
cgcccttcgg ctcggtctacca ataacgtggtc aacacgctga acagtgccgcgc aatatttgcctt 360
ttaaggtcag tggaaagaga attcggcagga tttaaaagat gtagcgtgaca aataataagc 420
tagagg 426

<210> SEQ ID NO 276
<211> LENGTH: 392
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 276
aggtgattcct cgcggctcttg ctcggcagcgg aatctgagaag cgttaaaat attcacttga 60
cgccggacacatt tggtacacta aagggcaggg ggacgaccttc gcgggtcttt 120
agggcagagtt ggtgcgcttc tggaaatagta tcgacccggt cctgggtgtaa atcggcactg 180
<210> SEQ ID NO 277
<211> LENGTH: 392
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 277
aatatcagg ccaaactgcac gttgcaaaac tgaacatcga tcacaaaccct gcacactgcgc 240
cgaatatgc catcctgcgt atccgcactc tgcgcgtgtt ccacaaagct gaagtgcgcg 300
cacaacaaagt gggtgcctgc ttcacaggtc aagttgaaga gacccctgcgcc gattataaag 360
atgcagcatga caaaaataaa gctagaggtca cc 392

<210> SEQ ID NO 278
<211> LENGTH: 528
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 278
gaattccgg gcttttttga ctggtctgtaa tgaatcctag gagggtctgc atatgaaat 60
aaaaacaggt gcgcgtcact gctgattact cgcattaacg acagtgatgt tttcgccttc 120
ggttcgcgc caaaaactcag aagccgcctc gcgcgtcctg gttgacgccc aatactagaa 180
cgaaaaattt attccctcga ctgacgacag ttttgcacgc gatgctacta aacgcagcgc 240
ggctgctgct gcgtttctcg ggcctagctgt gcgggcctgc tcgaaatacg tcggccgcgt 300
tcgcgtagac atcgctgcag aataatcccc gcaacctgcac gttgcaaaac tgaacatcga 360
tacaaaccct gcacactgcc cgaatattgc catcctgcgt atccgcactc tgcgcgtgtt 420
cacaacaaagt gggtgcctgc ttcacaggtc aagttgaaga gacccctgcgcc gattataaag 480
gacccctgcgcc gattataaag atgcagcatga caaaaataaa gctagaggtca cc 528

<210> SEQ ID NO 279
<211> LENGTH: 514
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 279
ccataaagg ctttttttggc tagccaggag ccctgcata tgaataaaaa acaggtgcac 60
gcatcctgcc atatacgcga ttaacagcga tgtggttttt cgcgctgcgct gcggccaaa 120
tcattgaag gcccctggcgc gcgtggtctg acgcgatc tagaagccgt aaattatcgc 180
-continued

acctgactga cgcagctttt gacacggatg tacattaagsac ggaaggggcgc atcttcgctcg 240
atttttgccg agaatggtgag ctgcgtctggca aataatcggcc ccgatttctg gatgaactcg 300
tgacgaaata tcagggcaaa ctggacgttg caaactgaa caatacataa aacccctgcca 360
tggtgctggaa atatgtgatc ctggtgtcccc cagatcgcgt gttgctcataaa aacggtgaaag 420
tggcggcact caaagcggtt gcaagtccatt aagttgctgtt caagagagaagctggtgatt 480
atataagata cggagcagaca taataagctata gagg 514

<210> SEQ ID NO 280
<211> LENGTH: 494
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 280

agggagtctc gcattagaaa ataaaaacag gtgcagccat cctccgatta tcggcattaa 60
cgcagatgat gtttttccgcc ttcgctcctcg ccaaataatc cgaagccgcc ctgaggggct 120
cggtggccgc cgaatttctg aagcataaaa ttatctacct gacttgacgc agttttgaca 180
cggagatgact caaagcgagt ccgggcatcc cgtgcaattt ttcggcgagag tgggtgcccgc 240
cgtgcaaata gtcagccccc atttgatcct aatctgctgta cgaatatacg gcggacattg 300
cgatttcgaa actcaaatcgc gatcaaaacc cttggactcgc gcggatataa ggcatccttg 360
gttccggac ccctgctgctg ccgaaaccgc gtgaagcttg gcggaccccaaa tgcggctgcac 420
tgctcuaaggg tcgggtgaaa gagaacctgg cggtattaaa agatgacgat gacaataa 480
aagctaggg tacc 494

<210> SEQ ID NO 281
<211> LENGTH: 494
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 281

agggagtctc gcattagaaa ataaaaacag gtgcagccat cctccgatta tcggcattaa 60
cgcagatgat gtttttccgcc ttcgctcctcg ccaaataatc cgaagccgcc ctgaggggct 120
cggtggccgc cgaatttctg aagcataaaa ttatctacct gacttgacgc agttttgaca 180
cggagatgact caaagcgagt ccgggcatcc cgtgcaattt ttcggcgagag tgggtgcccgc 240
cgtgcaaata gtcagccccc atttgatcct aatctgctgta cgaatatacg gcggacattg 300
cgatttcgaa actcaaatcgc gatcaaaacc cttggactcgc gcggatataa ggcatccttg 360
gttccggac ccctgctgctg ccgaaaccgc gtgaagcttg gcggaccccaaa tgcggctgcac 420
tgctcuaaggg tcgggtgaaa gagaacctgg cggtattaaa agatgacgat gacaataa 480
aagctaggg tacc 494

<210> SEQ ID NO 282
<211> LENGTH: 1521
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 282

gaatcatcgg gccttttcaag ctggtcgtaaa tgaaatccag gaggttcctcg atatggaaat 60
aaaaacaggt gcacgcatcc ctcgacttatc cgctattaaac gcagtaagcgtg tttaccgcct 120
gggtctggcg aaatcaagag aatattgcag ataaccgcct attaagccgg gataagcga 190
taacgttcctc gctgaaatcc gtaagaaatt cggagaaagat acggaaatct aattcaccgtt 240
tgacactcgg gataaactcg aagaaagaatt cccacggtct gccgaacagct gcagtggc 300
tgacattatc ttctgggcac aagacgcgttt tggtgcttac gcacatctg gccttggtgc 360
tgaatatac ggccgccagc cgttcaagca caagctgtat cctttaacct ggattgcctg 420
acgttaaac gcctaatct gttcttaccct gttgacgttg gaaagctttat cgcctgatta 480
taacaagct ctggcatcgg acaccgcaaaa aacccggaga gattcaccg cgtgggtat 540
agaaagct ggttcgcaac aagttccccgt gattgtaaat atggcagcaac acatgacac 600
cctgcgcttg tcctggtctc acgggggtta tcggttcag tattgaaacgc gcgaatccgc 660
catttaaagct gtgggcttggt ataagctgct gcggcaagcg ggtctgcanct tcctggtgta 720
cctggattaa aacaaacggt tattggtca gaccagttacc tccatcgcag aatgctgcct 780
taataaagg gccgatcctg gcggcgaggg gcggagttcg gctagtctca acacttgacac 840
cactcaagtgg aataatgtta taagcgtact gccgaccttc aaggtcctcacc catccaaacc 900
gttccgccgc tggcatccgct cggagttatt gcggccaggt ggtaaagatcg gagaagcgaa 960
agactcggcct tcattcctgta tggagtctg gatggtgatt tataaacaagcg 1020
accgcttcg ctgggcttgctg tgaattctca cggagtaagc tggcggagat atccacctt 1080
tgctgcgcacct atggaaaaac cccagtcgcgg tttctgtgat gctggctgta tcgaagcccg 1140
cctgaccgcg tggctgcgct gcggcaatag cgaaatctaa atattacccc tgaactcggc 1200
caggcttcg cctgtaatgc tcaaaaagccg gggggagtce tcctgctgatt tcggttccga 1260
gttctgctg gctttttcgc cgtttgctgg gatttctgccg tttctggcacc gaaaactgatga 1320
gggtggcgacct acgagttgca aacgtgacac cggcataacc cctgctcccgt cgccgaaata 1380
tgcgatctcg ggacactcgg ttccgcaatt cggcggaaaac ggtggtgggg gcggacacca 1440
agttgcttcg cttgctgctg tggcggaaaag ggtgtaatcg gggaaaccgaa 1500
tgcaaacat taagctcagag g 1521

<210> SEQ ID NO 283
<211> LENGTH: 1507
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 283

ccataccctt ttttttgggc taaacgaggg ccctgcctat gaaatattaa acaggtgccac 60
gcatctgcg atatttgcca ttaaagcaga tgattttttgc ccggccgctgct ccgcaaaaaa 120
tcggaaggg ctaatctctgcc aatccgcatcc gccgatcgcct atgggtccctg 180
aagttggcgg gcgaatccag gaattacccg cacoacggtgg atatgttccc gctggcggaa 240
aatgggaaga gaattcoca caggtgctgg ccaatgctgcga cggccgctgct ccgatcgcctg 300
gggcacaacg ccgctgttgg gcagtcggtgc atatggccct gttggtctgg gtaaaccgg 360
acaaagcgtt ccggaccaag ctgtatcctg ttacctggga tgcogtaacgt tacaacggsa 420
tagctgtgg tcacccgagc gcgtgtcagag cgtagtctg gatttatatac aagaatgtgc 480
tgcgaaacc gcgaaaaccc tgggaagaga tcccgccgct gcgataaaga atgaaagcga 540
aatgttacag egcgactagtc ttcacactgc aagacccgta ttccaccttg cggctgttgg 600
tctgctacgg gggtatcgcg ttcaagtatg aaacgacca gtcacagtctt aaagaagcgag 660
ggcggttatac ggcctgtgcca aagacgggct cgaccttacct gcgtaacctg attaanaaca 720
aacacatgaa tgcagacacc gattactcca tcgcagaaga tcgcttttaaat aaaggcagaa 780
cagcgcgtac catcaacgcgc cgctgacgat tgcaccaact cgcacaccgc aagatggaat 840
agtgttactc ggtatcggc acctttcagg tgtacccgct cccagccttc ctggtgcgtgc 900
tgacgocag tataagcgca gcacctccga acaaaagct gcggaaaaag ttccctcgaaa 960
actatcgtc gacgtagaag ggtgtggaag cgcggtaaat aagcacaacgc ctggggtgccg 1020
tgcgcgtcag gatccatgag gcagagtgtcg cgaagaatcc acgatgttgc gcaccctcgtg 1080
aaaacgccca gtcgcttccc ttgctgaccc tgcgtatcga agcgccctcg cagcctcctgg 1140	
tcgacogcga acatcagcag gataaataa ttccccctagac gcagcagact tgtgccagaat 1200
agttactctca aggacggggg ggcagtcctg tcgattttcg ggacaggttg tgcggtcgt 1260
gcgaattagt ccgccaagct ctgtctgaaa tgcgtagcgt gcctctgga aatcagacgg 1320	
tgcgaacnaa cacatagctt cccacccgct gcctggacgc ctaataatgc atcgtgcgtta 1380
tcccgacgct tcggcttgctg cccacggatt gcggtggccg ccaccagctg ggtgtcactg 1440
cataaagctca gttggaagag aaccttgccgg attatattga tgcagtagac aaataaaag 1500
cataagg 1507

<210> SEQ ID NO 284
<211> LENGTH: 1476
<220> ORGANISM: Artificial Sequence
<223> OTHER INFORMATION: Synthetic nucleic acid sequence
<225> SEQ: 284

agagggttct gcataagaaa ataaaaacag gtgcgaagcat ctctgcttattctgctaataaa 60
cgacccggt gttttcgcgc tcggctctcg ccasaatcga aagaagttaaa atggttatct 120
ngatcctagg gataaagcg tataacggct tcgctgcaagt cgtagtcatg ttcgagaaga 180
aatccggagtaa ataaaatccgg tggagcatac ggcataaat ggaaagttaaa tccacagcgg 240
ntggccagaac tgtggttcctg acgtgcatttt ttttcttgcc gcacgaccc agccccactgc 300
agcgactactg ccggcgtgcgt gcgatagtaa atcccggcaca acggagttcg gacagctgct 360
attctgcttcg tcggatgtgcc ttgacgttga acggcgaagct gatgtcatac gcgatcctgc 420	
tgtgacggtt atgcggtcag ataatcacaag atcgtgcggcc gcaccgcccg cccacctggg 480
aagagatccg gcggctgtggt aagagatcgg aagcycnagta atgacgacgc atgatactta 540
acgtgcagina acgtgtcacttc acgtgcgctc cgtagtctga tgtagctgctgtaa 600
agtagaaaa gcgcagtttc tagcataaag acgtgccttg gcctctgcc gcgcagcttc 660
eggctggcag ccggactgtgg gatgggttatta aaaaacaaca cgtatgcca gcacaagatt 720
acccagcgcc agaagtcggcct tttatataggg gcggacaccgc gcagccgtgct gacggccctg 780
<210> SEQ ID NO 285
<211> LENGTH: 1476
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<200> SEQUENCE: 285

agggaggttc gcataacagaa aacagcagaag taaaattatg tgaacgcgtag ctggccgactt 840
tcaaggttca aacacagaa acgcttgctg gctgcttgag gcaaggtatt aacacagcaa 900
gtcgagaac agagctggcc caaaggttccta tggaaacacta tcttctgactc gataaggttc 960
tggagaggt tataaaagc gaacgctggcttg gcggctggct gccgaagtct tagggagaggt 1020
tagttgcaag atagttcagtt atggcggcga ccattggaaaca ccgcaagcaggt gttctcttgt 1080
atggctcgtg tattcgaaggg ccgtgctggat cctgcgttcca gcggcaaatct agaggtcata 1140
aaattatca cctgaacgtg gactcgttgtag acacacagcagctgcaagggcaga 1200
tctctcagaa tttcctgctgag cgagtggtcgc gcgtcgtcaga atggtcgcagc cgggtctgttg 1260
agtaagtcgc ttcgcaaatct cagggcaaaag tgcagcctg gacaggtcagc acagttctgagt 1320
acacggcgcac taggctccct gcggagctgc gcttgctgctg caggtctaaaa ccctctctggttct 1380
aaggtgcatg ggctggcaacct aagtaaggttg cactgtctaaag aggctgagttttaaagagagaaaaacc 1440
tggcgtgtata aatcaggtcag gatgctaatc antaag 1476

<210> SEQ ID NO 286
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

agggaggttgc gcataagaa ataaaggggt tggcaagcgtc cctcgctaat cgcgctaat 60
cagagctgag aacgcagcgag ccataagaggtc aacacagcgg gaaagtaactg ctgtatatc 120
gagatggcag cagtaagaggtc ataaagatggc ggtgagaatg cggtaagaaaa tctcgagaaa 180
attacggat taaaagcgacgcc gatagacgcc gggagaaagaatgtcaagcgg 240
ttcgcaagctc tcctggctgc gctagctacg tgtgagctcag ctgtgagctctcgctg 300
aagctgctcag gttctggctg cgtgaggtcag ccgctgagcctg caggtgattgtgcgc 360
atctgctgatt ttcagaggtc atacagatgct cataacttcagc aacacgggaacc gacggtgtgtg 420
tggaagcggct attggcgcagt tataaaagag tcaagachtg cttctgtgca gacagcgg 480
aagaggtctg ggcagctgtag aagtaagaggtg tggagagggct gataaggctc gttgatct 540
aacgctggca caaagagcttc aacgctggcag ctgctggggtgt tctgctttgg 600
agatgagagag ccagctccgg cagcagctgag cggatgagctg ggcgaggaag 660
cgggtctgcg cttcgctggt gcagcgttatt gaaaaacaaaaatgtaagcctg gacaacggctt 720
acacggcagc agaggtgttcgcc ttaaaagagaag cagcagcagcagc gatgagcgtc 780
gggcctgcgtc caacagcttc aacagctgct gcggcttctgg ggaacgctggctg 840
tcaaggttca aacaczagagac cgctgcttgag gcagcatatt aacgcgca 900
gtcgacagaa acgctgttccag aaagagttcc gtaaaggactc gataagggtc 960
tggagaggt tataaaagac gacagcgcggg gctgctggctgc gcttgagtc tccggagagg 1020
agttggcagaa atagtcagttt atggcgtgca cctgtggagaaacgcgcagc gttctctctgtg 1080
atggctgcgtg tattcgaaggg ccgtgctggat cctgcgttcca gcggcaaatct agaggtcata 1140
aaattatca cctgaacgtg gactcgttgtag acacacagcagctgcaagggcaga 1200
tctctcagaa tttcctgctgag cgagtggtcgc gcgtcgtcaga atggtcgcagc cgggtctgttg 1260
<210> SEQ ID NO 286
<211> LENGTH: 383
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 286

Met Phe Glu Pro Met Glu Leu Thr Asn Asp Ala Val Ile Lys Val Ile
1    5      10     15

Gly Val Gly Gly Gly Gly Gly Asn Ala Val Glu His Met Val Arg Glu
20   25     30

Arg Ile Glu Gly Val Glu Phe Phe Ala Val Asn Thr Asp Ala Gln Ala
35   40     45

Leu Arg Lys Thr Ala Val Gly Gin Thr Ile Gin Ile Gly Ser Gly Ile
50   55     60

Thr Lys Gly Leu Gly Ala Gly Ala Asn Pro Glu Val Gly Arg Asn Ala
65   70     75   80

Ala Asp Glu Asp Arg Asp Ala Leu Arg Ala Ala Leu Glu Gly Ala Asp
85   90     95

Met Val Phe Ile Ala Ala Gly Met Gly Gly Gly Thr Gly Thr Gly Ala
100  105    110

Ala Pro Val Val Ala Glu Val Ala Lys Asp Leu Gly Ile Leu Thr Val
115  120    125

Ala Val Thr Lys Pro Phe Asn Phe Glu Gly Lys Lys Arg Met Ala
130  135    140

Phe Ala Glu Gln Gly Ile Thr Glu Leu Ser Lys His Val Asp Ser Leu
145  150    155   160

Ile Thr Ile Pro Asn Asp Lys Leu Leu Gly Val Gly Arg Gly Ile
165  170    175

Ser Leu Leu Asp Ala Phe Gly Ala Ala Asn Asp Val Leu Lys Gly Ala
180  185    190

Val Gln Gly Ile Ala Glu Leu Ile Thr Arg Pro Gly Leu Met Asn Val
195  200    205

Asp Phe Ala Asp Val Arg Thr Val Met Ser Gly Met Gly Tyr Ala Met
210  215    220

Met Gly Ser Gly Val Ala Ser Gly Glu Asp Arg Ala Glu Glu Ala Ala
225  230    235   240

Glu Met Ala Ile Ser Ser Pro Leu Leu Glu Asp Ile Asp Leu Ser Gly
245  250    255

Ala Arg Gly Val Leu Val Asn Ile Thr Ala Gly Phe Asp Leu Arg Leu
260  265    270

Asp Glu Phe Glu Thr Val Gly Asn Thr Ile Arg Ala Phe Ala Ser Asp
275  280    285

Asn Ala Thr Val Val Ile Gly Thr Ser Leu Asp Pro Asp Met Asn Asp
290  295    300

Glu Leu Arg Val Thr Val Val Thr Gly Ile Gly Met Asp Lys Arg
305  310    315

-continued

Pro Glu Ile Thr Leu Val Thr Asn Lys Gln Val Gln Gln Pro Val Met
325 330 335
Asp Arg Tyr Gln Gln His Gly Met Ala Pro Leu Thr Gln Glu Gln Lys
340 345 350
Pro Val Ala Lys Val Val Asp Asn Ala Pro Gin Thr Ala Lys Glu
355 360 365
Pro Asp Tyr Leu Asp Ile Pro Ala Phe Leu Arg Lys Gln Ala Asp
370 375 380

<210> SEQ ID NO: 287
<211> LENGTH: 383
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 287
Met Phe Glu Pro Met Glu Leu Thr Asn Asp Ala Val Ile Lys Val 1
5 10 15
Gly Val Gly Gly Gly Gly Gly Asn Ala Val Glu His Met Val Arg Glu
20 25 30
Arg Ile Glu Gly Val Phe Phe Ala Val Asn Thr Asp Ala Gln Ala
35 40 45
Leu Arg Lys Thr Ala Val Gly Gin Thr Ile Gin Ile Gly Ser Gly Ile
50 55 60
Thr Lys Gly Leu Gly Ala Gly Ala Asn Leu Glu Val Gly Arg Asn Ala
65 70 75 80
Ala Asp Glu Asp Arg Asp Ala Leu Arg Ala Ala Leu Glu Gly Ala Asp
85 90 95
Met Val Phe Ile Ala Ala Gly Met Gly Gly Gly Thr Gly Thr Gly Ala
100 105 110
Ala Pro Val Val Ala Glu Val Ala Lys Leu Asp Gly Ile Leu Thr Val
115 120 125
Ala Val Thr Lys Pro Phe Asn Phe Glu Gly Lys Arg Met Ala
130 135 140
Phe Ala Glu Gln Gly Ile Thr Glu Leu Ser Lys His Val Asp Ser Leu
145 150 155 160
Ile Thr Ile Pro Asn Asp Lys Leu Leu Lys Val Leu Gly Arg Gly Ile
165 170 175
Ser Leu Leu Asp Ala Phe Gly Ala Ala Asn Asp Val Leu Lys Gly Ala
180 185 190
Val Glu Gly Ile Ala Glu Leu Ile Thr Arg Pro Gly Leu Met Asn Val
195 200 205
Asp Phe Ala Asp Val Arg Thr Val Met Ser Glu Met Gly Tyr Ala Met
210 215 220
Met Gly Ser Gly Val Ala Ser Gly Glu Asp Arg Ala Glu Glu Ala Ala
225 230 235 240
Glu Met Ala Ile Ser Ser Pro Leu Leu Gly Asp Ile Asp Leu Ser Gly
245 250 255
Ala Arg Gly Val Leu Val Asn Ile Thr Ala Gly Phe Asp Leu Arg Leu
260 265 270
Asp Glu Phe Glu Thr Val Gly Asn Thr Ile Arg Ala Phe Ala Ser Asp
<table>
<thead>
<tr>
<th></th>
<th>270</th>
<th>280</th>
<th>290</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aaa Ala Thr Val Val Ile Gly Thr Ser Leu Asp Pro Asp Met Asn Asp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>295</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Glu Leu Arg Val Thr Val Val Ala Thr Gly Ile Gly Met Asp Lys Arg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>310</td>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td>Pro Glu Ile Thr Leu Val Thr Asn Lys Gin Val Gin Gin Pro Val Met</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Asp Arg Tyr Gin Gin His Gly Met Ala Pro Leu Thr Gin Glu Gin Lys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Pro Val Ala Lys Val Val Asn Asp Asn Ala Pro Gin Thr Ala Lys Glu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Pro Asp Tyr Leu Asp Ile Pro Ala Phe Leu Arg Lys Gin Ala Asp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 280
<211> LENGTH: 446
<212> TYPE: FMT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 288

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Pro Val Leu Glu Asn Arg Ala Ala Gin Gly Asp Ile Thr Ala Pro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Gly Gly Ala Arg Arg Leu Thr Gly Asp Gin Thr Ala Ala Leu Arg Asp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Ser Leu Ser Arg Gly Leu Leu Val Lys Thr Ala Ala Lys Gin Ala Ile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Gly Met Gly Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
</tr>
<tr>
<td>Ala Gly Gly Phe Phe Phe Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td>Tyr Thr His Tyr Ala Leu Asn Lys Thr Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
</tr>
<tr>
<td>Thr Asp Ser Ala Ala Ser Ala Thr Ala Thr Thr Thr Thr Thr Thr Thr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
</tr>
<tr>
<td>Tyr Aaa Ala Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Ile Leu Glu Met Ala Lys Ala Ala Gly Leu Ala Thr Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>115</td>
</tr>
<tr>
<td>Thr Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>130</td>
</tr>
<tr>
<td>Thr Ser Arg Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>145</td>
</tr>
<tr>
<td>Gly Aaa Ala Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>165</td>
</tr>
<tr>
<td>Leu Aaa Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>180</td>
</tr>
<tr>
<td>Ala Glu Thr Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>195</td>
</tr>
<tr>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>210</td>
</tr>
<tr>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td></td>
</tr>
<tr>
<td></td>
<td>225</td>
</tr>
<tr>
<td>Asn Ser Val Thr Glu Ala Asn Gin Gin Lys Thr Leu Leu Gly Lys Thr</td>
<td></td>
</tr>
</tbody>
</table>
|    |    |    |    |    | Phe
<table>
<thead>
<tr>
<th>245</th>
<th>250</th>
<th>255</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>Asp</td>
<td>Gly</td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>His</td>
<td>Gly</td>
<td>Asn</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Ser</td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Glu</td>
<td>Leu</td>
<td>Leu</td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Ala</td>
<td>Ser</td>
<td>Ile</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Thr</td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Gly</td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>His</td>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Asn</td>
<td>Ser</td>
<td>Glu</td>
</tr>
<tr>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Tyr</td>
</tr>
<tr>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Thr</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td>440</td>
<td>445</td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 289
<211> LENGTH: 494
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: <223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 289

Met| Ser| Arg| Pro| Arg| Leu| Ile| Ala| Leu| Phe| Leu| Phe| Phe| Asn| Val |
| 1 | 5 | 10 | 15 |
| Phe| Val| His| Gly| Glu| Asn| Lys| Val| Lys| Gin| Ser| Thr| Ile| Ala| Leu| Ala |
| 20 | 25 | 30 |
| Leu| Leu| Pro| Leu| Leu| Phe| Thr| Pro| Val| Thr| Lys| Ala| Arg| Thr| Pro| Glu |
| 35 | 40 | 45 |
| Met| Pro| Val| Leu| Glu| Asn| Arg| Ala| Ala| Gin| Gly| Asp| Ile| Thr| Ala| Pro |
| 50 | 55 | 60 |
| Gly| Gly| Ala| Arg| Arg| Leu| Thr| Gly| Asp| Gin| Thr| Ala| Leu| Arg| Asp |
| 65 | 70 | 75 | 80 |
| Ser| Leu| Ser| Asp| Lys| Pro| Ala| Lys| Arg| Lys| Ala| Ala| Leu| Leu| Ile| Gly |
| 85 | 90 | 95 |
| Gly| Met| Gly| Asp| Ser| Gln| Thr| Ala| Ala| Arg| Asn| Tyr| Ala| Glu| Gly |
| 100 |105 | 110 |
| Ala| Gly| Gly| Phe| Phe| Lys| Gly| Ile| Asp| Ala| Leu| Phe| Pro| Thr| Gln |
| 115 |120 | 125 |
| Tyr| Thr| His| Tyr| Ala| Leu| Asn| Lys| Lys| Thr| Gly| Leu| Pro| Asp| Tyr| Val |
| 130 |135 | 140 |
| Thr| Asp| Ser| Ala| Ala| Ser| Ala| Thr| Ala| Trp| Ser| Thr| Gly| Val| Lys| Thr |
Tyr Asn Gly Ala Leu Gly Val Asp Ile His Glu Lys Asp His Pro Thr
145 150 155 160
Ile Leu Glu Met Ala Lys Ala Ala Gly Leu Ala Thr Gly Asn Val Ser
165 170 175
Thr Ala Glu Leu Gln Asp Ala Thr Pro Ala Ala Leu Val Ala His Val
180 185 190 195 200 205
Thr Ser Arg Lys Cys Tyr Gly Pro Ser Ala Thr Ser Glu Lys Cys Pro
210 215 220
Gly Asn Ala Leu Glu Lys Gly Gly Ser Ile Thr Glu Gln Leu
225 230 235 240
Leu Asn Ala Arg Ala Asp Val Thr Leu Gly Gly Gly Ala Lys Thr Phe
245 250 255
Ala Glu Thr Ala Thr Ala Gly Glu Trp Gin Gly Lys Thr Leu Arg Glu
260 265 270
Gln Ala Gin Ala Arg Gly Tyr Gin Leu Val Ser Asp Ala Ala Ser Leu
275 280 285
Asn Ser Val Thr Glu Ala Asn Gin Gin Lys Pro Leu Leu Gly Leu Phe
290 295 300
Ala Asp Gly Asn Met Pro Val Arg Trp Leu Gly Pro Lys Ala Thr Tyr
305 310 315 320
His Gly Asn Ile Asp Lys Pro Ala Val Thr Cys Thr Pro Asn Pro Gin
325 330 335
Arg Asn Asp Ser Val Pro Thr Leu Ala Gin Met Thr Asp Lys Ala Ile
340 345 350
Glu Leu Leu Ser Lys Asn Gin Lys Gly Phe Phe Leu Gin Val Glu Gly
355 360 365
Ala Ser Ile Asp Lys Gln Asp His Ala Asn Pro Cys Gly Gin Ile
370 375 380
Gly Glu Thr Val Asp Leu Asp Glu Ala Val Gin Arg Ala Leu Glu Phe
385 390 395 400
Ala Lys Lys Glu Gly Asn Thr Leu Val Ile Val Thr Ala Asp His Ala
405 410 415
His Ala Ser Gin Ile Val Ala Pro Asp Thr Lys Ala Pro Gly Leu Thr
420 425 430
Gln Ala Leu Asn Thr Lys Asp Ala Val Met Val Met Ser Tyr Gly
435 440 445
Asn Ser Glu Glu Asp Ser Gin His Thr Gly Ser Gin Leu Arg Ile
450 455 460
Ala Ala Tyr Gly Pro His Ala Asn Val Gly Leu Thr Asp Gln
465 470 475 480
Thr Asp Leu Phe Tyr Thr Met Lys Ala Ala Leu Gly Leu Lys
485 490

<210> SEQ ID NO 290
<211> LENGTH: 468
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic chimeric amino acid sequence

<400> SEQUENCE: 290

Met Asn Leu Gly Asn Arg Leu Phe Ile Leu Ile Ala Val Leu Leu Pro
<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Leu Ala Val Leu Leu Leu Met Pro Val Leu Glu Asn Arg Ala Ala Gln</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Gly Asp Ile Thr Ala Pro Gly Gly Ala Arg Arg Leu Thr Gly Asp Gln</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Thr Ala Ala Leu Arg Asp Ser Leu Ser Asp Lys Pro Ala Lys Asn Ile</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Ile Leu Leu Ile Gly Asp Gly Met Gly Asp Ser Glu Ile Thr Ala Ala</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Arg Asn Tyr Ala Glu Gly Ala Gly Gly Phe Phe Lys Gly Ile Asp Ala</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Leu Pro Leu Thr Gly Gln Tyr Thr His Tyr Ala Leu Asn Lys Lys Thr</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Gly Lys Pro Asp Tyr Val Thr Asp Ser Ala Ala Ser Ala Thr Ala Trp</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>Ser Thr Gly Val Lys Thr Tyr Asn Gly Ala Leu Gly Val Asp Ile His</td>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Glu Lys Asp His Pro Thr Ile Leu Glu Met Ala Lys Ala Ala Gly Leu</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>Ala Thr Gly Asn Val Ser Thr Ala Glu Leu Gin Asp Ala Thr Pro Ala</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>Ala Leu Val Ala His Val Thr Ser Arg Lys Cys Tyr Gly Pro Ser Ala</td>
<td>180</td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>Thr Ser Glu Lys Cys Pro Gly Asn Ala Leu Glu Lys Gly Gly Lys Gly</td>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Ser Ile Thr Glu Gin Leu Leu Asn Ala Arg Ala Asp Val Thr Leu Gly</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>Gly Gly Ala Lys Thr Phe Ala Glu Thr Ala Thr Ala Gly Glu Trp Gin</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td></td>
<td>Gly Lys Thr Leu Arg Glu Gin Ala Glu Ala Arg Gly Tyr Gin Leu Val</td>
<td>245</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>Ser Asp Ala Ala Ser Leu Asn Ser Val Thr Glu Ala Asn Gin Gin Lys</td>
<td>260</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>Pro Leu Leu Gly Leu Phe Ala Asp Gly Asn Met Pro Val Arg Trp Leu</td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Gly Pro Lys Ala Thr Tyr His Gly Asn Ile Asp Lys Pro Ala Val Thr</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>Cys Thr Pro Asn Pro Gin Arg Asp Ser Val Pro Thr Leu Ala Gin</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>Met Thr Asp Lys Ala Ile Glu Leu Leu Ser Lys Asn Glu Lys Gly Phe</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td></td>
<td>Phe Leu Gin Val Glu Gly Ala Ser Ile Asp Lys Gin Asp His Ala Ala</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td></td>
<td>Asn Pro Cys Gly Gin Ile Gly Glu Thr Val Asp Leu Asp Gln Ala Val</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>Gin Arg Ala Leu Glu Phe Ala Lys Glu Gly Asn Thr Leu Val Ile</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td></td>
<td>Val Thr Ala Asp His Ala His Ala Ser Gin Ile Val Ala Pro Asp Thr</td>
<td>385</td>
<td>390</td>
</tr>
<tr>
<td></td>
<td>Lys Ala Pro Gly Leu Thr Gin Ala Leu Asn Thr Lys Asp Gly Ala Val</td>
<td>405</td>
<td>410</td>
</tr>
</tbody>
</table>
Met Val Met Ser Tyr Gly Asn Ser Glu Glu Asp Ser Gin Glu His Thr
420 425 430
Gly Ser Gln Leu Arg Ile Ala Ala Tyr Gly Pro His Ala Ala Asn Val
435 440 445
Val Gly Leu Thr Asp Gin Thr Asp Leu Phe Tyr Thr Met Lys Ala Ala
450 455 460
Leu Gly Leu Lys
465

<210> SEQ ID NO 291
<211> LENGTH: 424
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 291
Met Ala Thr Thr His Ala Gin Gly His Pro Pro Val Leu Gly Asn Asp
1  5 10  15
Thr Leu Arg Glu His Tyr Asp Tyr Val Gly Lys Leu Ala Gly Arg Leu
20  25 30
Arg Asp Pro Pro Glu Gly Ser Thr Leu Ile Thr Thr Ile Leu Phe Leu
35  40 45
Val Thr Cys Ser Phe Ile Val Leu Glu Asn Leu Met Val Leu Ile Ala
50  55  60
Ile Trp Lys Asn Asn Lys Phe His Asn Arg Met Tyr Phe Phe Ile Gly
65  70  75  80
Asn Leu Ala Leu Cys Asp Leu Leu Ala Gly Ile Ala Tyr Lys Val Asn
85  90  95
Ile Leu Met Ser Gly Arg Lys Thr Phe Ser Leu Ser Pro Thr Val Trp
100 105 110
Phe Leu Arg Glu Gly Ser Met Phe Val Ala Gly Ala Ser Thr Cys
115 120 125
Ser Leu Leu Ala Ile Ala Glu His Leu Thr Met Ile Lys Met
130 135 140
Arg Pro Tyr Asp Ala Asn Lys His Arg Val Phe Leu Leu Ile Gly
145 150 155 160
Met Cys Trp Leu Ile Ala Phe Ser Leu Gly Ala Leu Pro Ile Leu Gly
165 170 175
Trp Asn Cys Leu Glu Asn Phe Pro Asp Cys Ser Thr Ile Leu Pro Leu
180 185 190
Tyr Ser Lys Tyr Ile Ala Phe Leu Ile Ser Ile Phe Ile Ala Ile
195 200 205
Leu Val Thr Ile Val Ile Leu Tyr Ala Arg Ile Tyr Phe Leu Val Lys
210 215 220
Ser Ser Ser Arg Arg Val Ala Asn His Asn Ser Glu Arg Ser Met Ala
225 230 235 240
Leu Leu Arg Thr Val Val Val Ser Val Phe Ile Ala Cys Trp
245 250 255
Ser Pro Leu Phe Leu Leu Ile Asp Val Ala Cys Arg Ala Lys
260 265 270
Glu Cys Ser Ile Leu Phe Lys Ser Gln Trp Phe Ile Met Leu Ala Val
275 280 285
Leu Asn Ser Ala Met Asn Pro Val Ile Tyr Thr Leu Ala Ser Lys Glu
290     295     300
Met Arg Arg Ala Phe Phe Arg Leu Val Cys Gly Cys Leu Val Lys Gly
305     310     315     320
Lys Gly Thr Glu Ala Ser Pro Met Glu Pro Ala Leu Asp Pro Ser Arg
325     330     335
Ser Lys Ser Ser Ser Ser Asn Ser Ser Ser His Ser Pro Lys Val
340     345     350
Lys Glu Asp Leu Pro His Val Ala Thr Ser Ser Cys Val Thr Asp Lys
355     360     365
Thr Arg Ser Leu Glu Asn Gly Val Leu Cys Lys Gly Asn Ser Ala
370     375     380
Asp Ile Gln His Ser Gly Gly Arg Ser Ser Leu Glu Gly Pro Arg Phe
385     390     395     400
Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg
405     410     415
Thr Gly His His His His His
420

<210> SEQ ID NO 292
<211> LENGTH: 413
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: UNSURE
<222> LOCATION: 164
<223> OTHER INFORMATION: Xaa = unknown amino acid
<400> SEQUENCE: 292
Met Gly Glu Pro Gly Asn Gly Ser Ala Phe Leu Leu Ala Pro Asn Gly
1     5     10     15
Ser His Ala Pro Asp His Asp Val Thr Glu Glu Arg Asp Glu Val Trp
20     25     30
Val Val Gly Met Gly Ile Val Met Ser Leu Ile Val Leu Ala Ile Val
35     40     45
Phe Gly Asn Val Leu Val Ile Thr Ala Ile Ala Lys Phe Glu Arg Leu
50     55     60
Gln Thr Val Thr Asn Tyr Phe Ile Thr Ser Leu Ala Cys Ala Asp Leu
65     70     75     80
Val Met Gly Leu Ala Val Pro Phe Gly Ala Ala His Ile Leu Met
85     90     95
Lys Met Trp Thr Phe Gly Asn Phe Trp Cys Glu Phe Trp Thr Ser Ile
100    105
Asp Val Leu Cys Val Thr Ala Ser Ile Glu Thr Leu Cys Val Ile Ala
115    120    125
Val Asp Arg Tyr Phe Ala Ile Thr Ser Pro Phe Lys Tyr Glu Ser Leu
130    135
Leu Thr Lys Asn Lys Ala Arg Val Ile Ile Leu Met Val Trp Ile Val
145    150    155    160
Ser Gly Leu Xaa Ser Phe Leu Pro Ile Glu Met His Thr Tyr Arg Ala
165    170    175
Thr His Glu Ala Ile Asn Cys Tyr Ala Asn Glu Thr Cys Cys Asp
180    185    190
-continued

```
Phe  Phe  Thr  Asn  Gln  Ala  Tyr  Ala  Ile  Ala  Ser  Ser  Ile  Val  Ser  Phe
   195             200             205
Tyr  Val  Pro  Leu  Val  Ile  Met  Val  Phe  Val  Tyr  Ser  Arg  Val  Phe  Gln
   210             215             220
Glu  Ala  Lys  Arg  Gln  Leu  Gln  Lys  Ile  Asp  Lys  Ser  Glu  Gly  Arg  Phe
   225             230             235             240
His  Val  Gln  Asn  Leu  Ser  Gln  Val  Glu  Gln  Arg  Gly  Arg  Thr  Gly  His
   245             250             255
Gly  Leu  Arg  Arg  Ser  Ser  Lys  Phe  Cys  Leu  Lys  Glu  His  Lys  Ala  Leu
   260             265             270
Lys  Thr  Leu  Gln  Ile  Ile  Met  Gly  Thr  Phe  Thr  Leu  Cys  Trp  Leu  Pro
   275             280             285
Phe  Phe  Ile  Val  Asn  Ile  Val  His  Val  Ile  Glu  Asp  Asn  Leu  Ile  Arg
   290             295             300
Lys  Glu  Val  Tyr  Ile  Leu  Leu  Asn  Trp  Ile  Gly  Tyr  Val  Asn  Ser  Gly
   305             310             315             320
Phe  Asn  Pro  Leu  Ile  Tyr  Cys  Arg  Ser  Pro  Asp  Phe  Arg  Ile  Ala  Phe
   325             330             335
Gln  Glu  Leu  Leu  Cys  Leu  Arg  Arg  Ser  Ser  Leu  Lys  Ala  Tyr  Gly  Asn
   340             345             350
Gly  Tyr  Ser  Ser  Asn  Gly  Arg  Thr  Gly  Glu  Glu  Ser  Gly  Tyr  His  Val
   355             360             365
Glu  Gln  Glu  Lys  Glu  Asn  Leu  Leu  Cys  Glu  Asp  Leu  Pro  Gly  Thr
   370             375             380
Glu  Asp  Phe  Val  Gly  His  Gln  Gly  Thr  Val  Pro  Ser  Asp  Ile  Asp
   385             390             395             400
Ser  Glu  Gly  Arg  Arg  Cys  Ser  Thr  Asn  Asp  Ser  Leu  Leu
   405             410

<210> SEQ ID NO 293
<211> LENGTH: 416
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 293

Met  Asp  Ser  Val  Cys  Pro  Gln  Gly  Lys  Tyr  Ile  His  Pro  Glu  Asn  Asn
   1               5              10              15
Ser  Ile  Cys  Cys  Thr  Lys  Cys  His  Lys  Gly  Thr  Tyr  Leu  Tyr  Asn  Asp
   20              25              30
Cys  Pro  Gly  Pro  Gly  Glu  Asp  Thr  Arg  Cys  Arg  Glu  Cys  Glu  Ser  Gly
   35              40              45
Ser  Phe  Thr  Ala  Ser  Glu  Asn  His  Leu  Arg  His  Cys  Leu  Ser  Cys  Ser
   50              55              60
Lys  Cys  Arg  Lys  Glu  Met  Gln  Glu  Val  Glu  Ile  Ser  Ser  Cys  Thr  Val
   65              70              75              80
Asp  Arg  Asp  Thr  Val  Cys  Glu  Cys  Arg  Lys  Asn  Gln  Tyr  Arg  His  Tyr
   85              90              95
Trp  Ser  Glu  Asn  Leu  Phe  Gln  Cys  Phe  Asn  Cys  Ser  Leu  Cys  Leu  Asn
   100             105             110
Gly  Thr  Val  His  Leu  Ser  Cys  Gln  Glu  Lys  Glu  Asn  Thr  Val  Cys  Thr
   115             120             125
```
-continued

Cys His Ala Gly Phe Phe Leu Arg Arg Asn Glu Cys Val Ser Cys Ser
130 135 140
Asn Cys Lys Lys Ser Leu Glu Cys Thr Lys Leu Cys Leu Pro Gln Ile
145 150 155 160
Glu Asn Val Lys Gly Thr Glu Asp Ser Gly Thr Thr Val Leu Leu Pro
165 170 175
Leu Val Ile Phe Phe Gly Leu Cys Leu Leu Ser Leu Leu Phe Ile Gly
180 185 190
Leu Met Tyr Arg Tyr Gln Arg Trp Lys Ser Lys Leu Tyr Ser Ile Val
195 200 205
Cys Gly Lys Ser Thr Pro Glu Lys Gly Gly Leu Glu Gly Thr Thr
210 215 220
Thr Lys Pro Leu Ala Pro Asn Pro Ser Phe Ser Pro Thr Pro Gly Phe
225 230 235 240
Thr Pro Thr Leu Gly Phe Ser Pro Val Pro Ser Thr Phe Thr Ser
245 250 255
Ser Ser Thr Tyr Thr Pro Gly Asp Cys Pro Asn Phe Ala Ala Pro Arg
260 265 270
Arg Glu Val Ala Pro Pro Tyr Gin Gly Ala Asp Pro Ile Leu Ala Thr
275 280 285
Ala Leu Ala Ser Asp Pro Ile Pro Asn Pro Leu Gin Lys Thr Glu Asp
290 295 300
Ser Ala His Lys Pro Gin Ser Leu Asp Thr Asp Asp Pro Ala Thr Leu
305 310 315 320
Tyr Ala Val Val Glu Asn Val Pro Pro Leu Arg Trp Lys Glu Phe Val
325 330 335
Arg Arg Leu Gly Leu Ser Asp His Glu Ile Asp Arg Leu Glu Leu Gln
340 345 350
Asn Gly Arg Cys Leu Arg Arg Glu Ala Glu Tyr Ser Met Leu Ala Thr Trp
355 360 365
Arg Arg Arg Thr Pro Arg Arg Glu Ala Thr Leu Glu Leu Gly Arg
370 375 380
Val Leu Arg Asp Met Asp Leu Gly Cys Leu Gly Asp Ile Glu Glu
385 390 395 400
Ala Leu Cys Gly Pro Ala Ala Leu Pro Pro Ala Pro Ser Leu Leu Arg
405 410 415

<210> SEQ ID NO 294
<211> LENGTH: 428
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 294
Met Gly Leu Val Pro His Leu Gly Asp Arg Glu Lys Arg Asp Ser Val
1 5 10 15
Cys Pro Gln Gly Lys Tyr Ile His Pro Glu Asn Asn Ser Ile Cys Cys
20 25 30
Thr Lys Cys His Lys Gly Thr Tyr Leu Tyr Asn Cys Pro Gly Pro
35 40 45
Gly Glu Asp Thr Asp Cys Arg Glu Cys Glu Ser Gly Ser Phe Thr Ala
50 55 60
Continued

Ser Glu Asn His Leu Arg His Cys Leu Ser Cys Ser Lys Cys Arg Lys
65 70 75 80
Glu Met Gly Glu Val Glu Ile Ser Ser Cys Thr Val Asp Arg Asp Thr
85 90
Val Cys Gly Cys Arg Lys Asn Glu Tyr Arg His Tyr Thr Ser Glu Asn
100 105 110
Leu Phe Glu Cys Phe Asn Cys Ser Leu Cys Leu Asn Gly Thr Val His
115 120 125
Leu Ser Cys Glu Lys Glu Asn Thr Val Cys Thr Cys His Ala Gly
130 135 140
Phe Phe Leu Arg Glu Asn Glu Cys Val Ser Cys Ser Asn Cys Lys Lys
145 150 155 160
Ser Leu Glu Cys Thr Lys Leu Cys Leu Pro Glu Ile Glu Asn Val Lys
165 170 175
Gly Thr Glu Asp Ser Gly Thr Thr Val Leu Leu Pro Leu Val Ile Phe
180 185 190
Phe Gly Leu Cys Leu Ser Leu Leu Leu Phe Ile Gly Leu Met Tyr Arg
195 200 205
Tyr Glu Arg Trp Lys Ser Lys Leu Tyr Ser Ile Val Cys Gly Lys Ser
210 215 220
Thr Pro Glu Lys Glu Leu Glu Gly Thr Thr Thr Lys Pro Leu
225 230 235 240
Ala Pro Asn Pro Ser Phe Ser Pro Thr Pro Gly Phe Thr Pro Thr Leu
245 250 255
Gly Phe Ser Pro Val Pro Ser Ser Thr Phe Thr Ser Ser Ser Ser Thr Tyr
260 265 270
Thr Pro Gly Asp Cys Pro Asn Phe Ala Ala Ala Pro Arg Arg Glu Val Ala
275 280 285
Pro Pro Tyr Glu Gly Ala Asp Pro Ile Leu Ala Thr Ala Leu Ala Ser
290 295 300
Asp Pro Ile Pro Asn Pro Leu Glu Glu Thr Trp Glu Asp Ser Ala His Lys
305 310 315 320
Pro Glu Ser Leu Asp Thr Asp Asp Pro Ala Thr Leu Tyr Ala Val Val
325 330 335
Glu Asn Val Pro Pro Leu Arg Trp Lys Glu Phe Val Arg Arg Leu Gly
340 345 350
Leu Ser Asp His Glu Ile Asp Arg Leu Glu Leu Glu Asn Gly Arg Cys
355 360 365
Leu Arg Glu Ala Glu Tyr Ser Met Leu Ala Thr Trp Arg Arg Arg Thr
370 375 380
Pro Arg Glu Ala Thr Leu Glu Leu Gly Arg Val Leu Arg Asp
395 399 400
Met Asp Leu Leu Gly Cys Leu Glu Asp Ile Glu Glu Ala Leu Cys Gly
405 410 415
Pro Ala Ala Leu Pro Pro Ala Pro Ser Leu Leu Arg
420 425

<210> SEQ ID NO 295
<211> LENGTH: 233
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 295

Met Ser Thr Glu Ser Met Ile Arg Asp Val Glu Leu Ala Glu Glu Ala
1  5  10  15

Leu Pro Lys Lys Thr Gly Gly Pro Gln Gly Ser Arg Arg Cys Leu Phe
20  25  30

Leu Ser Leu Phe Ser Phe Leu Ile Val Ala Gly Ala Thr Thr Leu Phe
35  40  45

Cys Leu Leu His Phe Gly Val Ile Gly Pro Gin Arg Glu Glu Phe Pro
50  55  60

Arg Asp Leu Ser Leu Ile Ser Pro Leu Ala Gin Ala Val Arg Ser Ser
65  70  75  80

Ser Arg Thr Pro Ser Asp Lys Pro Val Ala His Val Val Ala Asn Pro
85  90  95

Gln Ala Glu Gly Gln Leu Gln Trp Leu Asn Arg Arg Ala Asn Ala Leu
100 105 110

Leu Ala Asn Gly Val Glu Leu Arg Asp Asn Gin Leu Val Val Pro Ser
115 120 125

Glu Gly Leu Tyr Leu Ile Tyr Ser Gin Val Leu Phe Lys Gly Gin Gly
130 135 140

Cys Pro Ser Thr His Val Leu Thr His Thr Ile Ser Arg Ile Ala
145 150 155 160

Val Ser Tyr Gln Thr Lys Val Asn Leu Leu Ser Ala Ile Lys Ser Pro
165 170 175

Cys Gin Arg Glu Thr Pro Glu Gly Ala Glu Ala Lys Pro Trp Tyr Glu
180 185 190

Pro Ile Tyr Leu Gly Gly Val Phe Gin Leu Glu Lys Gly Asp Arg Leu
195 200 205

Ser Ala Glu Ile Asn Arg Pro Asp Tyr Leu Asp Phe Ala Glu Ser Gly
210 215 220

Gln Val Tyr Phe Gly Ile Ile Ala Leu
225 230

<210> SEQ ID NO 296
<211> LENGTH: 75
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 296

Met Asn Leu Gly Asn Arg Leu Phe Ile Leu Ile Ala Val Leu Leu Pro
1  5  10  15

Leu Ala Val Leu Leu Leu Ser Asp Ser Glu Cys Pro Leu Ser His
20  25  30

Asp Gly Tyr Cys Leu His Asp Gly Val Cys Met Tyr Ile Glu Ala Leu
35  40  45

Asp Lys Tyr Ala Cys Asn Cys Val Val Gly Tyr Ile Gly Glu Arg Cys
50  55  60

Gln Tyr Arg Asp Leu Lys Trp Trp Glu Leu Arg
65  70  75

<210> SEQ ID NO 297
<211> LENGTH: 518

<223> OTHER INFORMATION: Synthetic amino acid sequence

Met Asn Leu Gly Asn Arg Leu Phe Ile Leu Ile Ala Val Leu Leu Pro
1  5   10  15
Leu Ala Val Leu Leu Leu Ser Phe Thr Leu Ser Val Thr Val Gln Gln
20  25  30
Pro Gln Leu Thr Leu Thr Ala Ala Val Ile Gly Asp Gly Ala Pro Ala
35  40  45
Asn Gly Lys Thr Ala Ile Thr Val Gln Phe Thr Val Ala Asp Phe Glu
50  55  60
Gly Lys Pro Leu Ala Gly Gln Glu Val Ile Thr Thr Asn Asn Gly
65  70  75  80
Ala Leu Pro Asn Lys Ile Thr Glu Lys Thr Asp Ala Asn Gly Val Ala
85  90  95
Arg Ile Ala Leu Thr Asn Thr Asp Gly Val Thr Val Thr Ala
100 105 110
Glu Val Glu Gly Gln Arg Glu Ser Val Asp Thr His Phe Val Lys Gly
115 120 125
Thr Ile Ala Ala Asp Lys Ser Thr Leu Ala Ala Val Pro Thr Ser Ile
130 135 140
Ile Ala Asp Gly Leu Met Ala Ser Thr Ile Thr Leu Glu Leu Asp
145 150 155 160
Thr Tyr Gly Asp Pro Gln Ala Gly Ala Asn Val Ala Phe Asp Thr Thr
165 170 175
Leu Gly Asn Met Gly Val Ile Thr Asp His Asn Asp Gly Thr Tyr Ser
180 185 190
Ala Pro Leu Thr Ser Thr Thr Leu Gly Val Ala Thr Val Thr Val Lys
195 200 205
Val Asp Gly Ala Ala Phe Ser Val Pro Ser Val Thr Val Asn Phe Thr
210  215 220
Ala Asp Pro Ile Pro Asp Ala Gly Arg Ser Ser Phe Thr Val Ser Thr
225 230 235 240
Pro Asp Ile Leu Ala Asp Gly Thr Met Ser Ser Thr Leu Ser Phe Val
245 250 255
Pro Val Asp Lys Asn Gly His Phe Ile Ser Gly Met Gln Gly Leu Ser
260 265 270
Phe Thr Gln Asn Gly Val Pro Val Ser Ile Ser Pro Ile Thr Glu Gln
275 280 285
Pro Asp Ser Tyr Thr Ala Thr Val Val Gly Asn Ser Val Gly Asp Val
290 295 300
Thr Ile Thr Pro Gln Val Asp Thr Leu Ile Leu Ser Thr Leu Gln Lys
305 310 315 320
Lys Ile Ser Leu Phe Pro Val Pro Thr Leu Thr Gly Ile Leu Val Asn
325 330 335
Gly Gln Asn Phe Ala Thr Asp Lys Gly Phe Pro Lys Thr Ile Phe Lys
340 345 350
Asn Ala Thr Phe Gln Leu Gln Met Asp Asn Asp Val Ala Asn Asn Thr
355 360 365
-continued

<table>
<thead>
<tr>
<th>Glu Tyr Glu Trp Ser Ser Ser Phe Thr Pro Asn Val Ser Val Asn Asp</th>
<th>370</th>
<th>375</th>
<th>380</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu Gly Glu Val Thr Ile Thr Tyr Gln Thr Tyr Ser Glu Val Ala Val</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Thr Ala Lys Ser Lys Lys Phe Pro Ser Tyr Ser Val Ser Tyr Arg Phe</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Tyr Pro Asn Arg Trp Ile Tyr Asp Gly Gly Arg Ser Leu Val Ser Ser</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Leu Glu Ala Ser Arg Glu Cys Glu Gly Ser Asp Met Ser Ala Val Leu</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Glu Ser Ser Arg Ala Thr Asn Gly Thr Arg Ala Pro Asp Gly Thr Leu</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Trp Gly Glu Trp Gly Ser Leu Thr Ala Tyr Ser Ser Asp Trp Gln Ser</td>
<td>465</td>
<td>470</td>
<td>475</td>
</tr>
<tr>
<td>Gly Glu Tyr Trp Val Lys Thr Ser Thr Asp Phe Glu Thr Met Asn</td>
<td>485</td>
<td>490</td>
<td>495</td>
</tr>
<tr>
<td>Met Asp Thr Gly Ala Leu Gln Pro Gly Pro Ala Tyr Leu Ala Phe Pro</td>
<td>500</td>
<td>505</td>
<td>510</td>
</tr>
<tr>
<td>Leu Cys Ala Leu Ser Ile</td>
<td>515</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 298
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 298

Met Ser Arg Pro Arg Leu Ile Val Ala Leu Phe Leu Phe Asn Val
1   5   10  15

Phe Val His Gly Glu Asn Lys Val Lys Gln Ser Thr Ile Ala Leu Ala
20  25  30

Leu Leu Pro Leu Leu Phe Thr Pro Val Thr Lys Ala Arg Thr Pro Glu
35  40  45

Ser Arg
50

<210> SEQ ID NO 299
<211> LENGTH: 496
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 299

Met Ser Arg Pro Arg Leu Ile Val Ala Leu Phe Leu Phe Asn Val
1   5   10  15

Phe Val His Gly Glu Asn Lys Val Lys Gln Ser Thr Ile Ala Leu Ala
20  25  30

Leu Leu Pro Leu Leu Phe Thr Pro Val Thr Lys Ala Arg Thr Pro Glu
35  40  45

Met Pro Val Leu Glu Asn Arg Ala Ala Glu Gly Asp Ile Thr Ala Pro
50  55  60

Gly Gly Ala Arg Arg Leu Thr Gly Asp Gin Thr Ala Ala Leu Arg Asp
65  70  75  80
Ser Leu Ser Asp Lys Pro Ala Lys Asn Ile Ile Leu Leu Ile Gly Asp 95 90 95
Gly Met Gly Asp Ser Glu Ile Thr Ala Ala Arg Asn Tyr Ala Glu Gly 100 105 110
Ala Gly Gly Phe Phe Lys Gly Ile Asp Ala Leu Pro Leu Thr Gly Gln 115 120 125
Tyr Thr His Tyr Ala Leu Asn Lys Thr Gly Lys Pro Asp Tyr Val 130 135 140
Thr Asp Ser Ala Ala Ser Ala Thr Ala Trp Ser Thr Gly Val Lys Thr 145 150 155 160
Tyr Asn Gly Ala Leu Gly Val Asp Ile His Glu Lys Asp His Pro Thr 165 170 175
Ile Leu Glu Met Ala Lys Ala Ala Gly Leu Ala Thr Gly Asn Val Ser 180 185 190
Thr Ala Glu Leu Gln Asp Ala Thr Pro Ala Ala Leu Val Ala His Val 195 200 205
Thr Ser Arg Lys Cys Tyr Gly Pro Ser Ala Thr Ser Glu Lys Cys Pro 210 215 220
Gly Asn Ala Leu Glu Lys Gly Gly Lys Gly Ser Ile Thr Glu Gln Leu 225 230 235 240
Leu Asn Ala Arg Ala Asp Val Thr Leu Gly Gly Gly Ala Lys Thr Phe 245 250 255
Ala Glu Thr Ala Thr Ala Gly Glu Trp Gln Gly Lys Thr Leu Arg Glu 260 265 270
Gln Ala Gln Ala Arg Gly Tyr Glu Leu Val Ser Asp Ala Ala Ser Leu 275 280 285
Asn Ser Val Thr Glu Ala Asn Gln Glu Leu Pro Leu Leu Gly Leu Phe 290 295 300
Ala Asp Gly Asn Met Pro Val Arg Trp Leu Gly Pro Lys Ala Thr Tyr 305 310 315 320
His Gly Asn Ile Asp Lys Pro Ala Val Thr Cys Thr Pro Asn Pro Gln 325 330 335
Arg Asn Asp Ser Val Pro Thr Leu Ala Gln Met Thr Asp Lys Ala Ile 340 345 350
Glu Leu Leu Ser Lys Asn Glu Lys Gly Phe Phe Leu Gln Val Glu Gly 355 360 365
Ala Ser Ile Asp Lys Gln Asp His Ala Ala Asn Pro Cys Gly Gln Ile 370 375 380
Gly Glu Thr Val Asp Leu Asp Glu Ala Val Gln Arg Ala Leu Glu Phe 385 390 395 400
Ala Lys Lys Glu Gly Asn Thr Leu Val Ile Val Thr Ala Asp His Ala 405 410 415
His Ala Ser Gln Ile Val Ala Pro Asp Thr Lys Ala Pro Gly Leu Thr 420 425 430
Gln Ala Leu Asn Thr Lys Asp Gly Ala Val Met Val Met Ser Tyr Gly 435 440 445
Asn Ser Glu Glu Asp Ser Gln Glu His Thr Gly Ser Gln Leu Arg Ile 450 455 460
Ala Ala Tyr Gly Pro His Ala Ala Asn Val Val Gly Leu Thr Asp Gln 465 470 475 480
Thr Asp Leu Phe Tyr Thr Met Lys Ala Ala Leu Gly Leu Lys Ser Arg
486 490 495

<210> SEQ ID NO 300
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 300

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1   5   10   15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ser Arg
20  25  30

<210> SEQ ID NO 301
<211> LENGTH: 372
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 301

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5   10   15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20  25  30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35  40

Val Gly Lys Tyr Glu Gly Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50  55  60

His Pro Asp Lys Leu Glu Gly Lys Phe Pro Gin Val Ala Ala Thr Gly
65  70  75  80

Asp Gly Pro Asp Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85  90  95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125

Leu Ile Ala Tyr Pro Ile Ala Val Ala Leu Ser Leu Ile Tyr Asn
130 135 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160

Leu Asp Lys Glu Leu Lys Ala Lys Gly Ser Ala Leu Met Phe Asn
165 170 175

Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205

Val Asp Asn Ala Gly Ala Lys Leu Thr Phe Leu Val Asp Leu
210 215 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
-continued

Ala Trp Ser Arg Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
  260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
  275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
  290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
  305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Gly Glu Glu
  325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Glu Lys
  340 345 350
Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Trp Tyr Ala
  355 360 365
Val Arg Ser Arg
  370

<210> SEQ ID NO 302
<211> LENGTH: 133
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 302
Met Ile Glu Ala Arg Ser Arg Leu Glu Ser Asp Lys Ile Ile His Leu
  1  5  10  15
Thr Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile
  20  25  30
Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala
  35  40  45
Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gin Gly Lys Leu Thr Val
  50  55  60
Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly
  65  70  75  80
Ile Arg Gly Ile Pro Thr Leu Leu Phe Lys Asn Gly Glu Val Ala
  85  90  95
Ala Thr Lys Val Gly Ala Leu Ser Lys Gin Gly Leu Lys Phe Leu
 100 105 110
Asp Ala Asn Leu Ala Leu Glu Asp Tyr Lys Asp His Asp Gly Asp Tyr
 115 120 125
Lys Asp His Asp Arg
 130

<210> SEQ ID NO 303
<211> LENGTH: 767
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 303
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
  1  5  10  15
Thr Met Met Phe Ser Ala Ser Leu Ala Lys Ile Glu Glu Gly Lys

Thr Met Met Phe Ser Ala Ser Leu Ala Lys Ile Glu Glu Gly Lys

Thr Met Met Phe Ser Ala Ser Leu Ala Lys Ile Glu Glu Gly Lys
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Glu Val Ala Ala Thr Gly
65 70 75 80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asn Pro Asp Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Gly Lys Ser Ala Leu Met Phe Asn
165 170 175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Glu Lys
340 345 350
Gly Glu Ile Met Pro Asn Ile Pro Gln Met Ser Ala Phe Thr Tyr Ala
355 360 365
Val Leu Ile Glu Ala Arg Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn
370 375 380
Ser Asp Leu Asp Val Asn Thr Ile Tyr Ser Lys Val Leu Val Thr
385 390 395 400
 Ala Ile Tyr Leu Ala Leu Phe Val Val Gly Thr Val Gly Asn Ser Val
405 410 415
Thr Ala Phe Thr Leu Ala Arg Lys Lys Ser Leu Gln Ser Leu Gln Ser
420 425 430
Thr Val His Tyr His Leu Gly Ser Leu Ala Leu Ser Asp Leu Leu Ile 435 440 445
Leu Leu Leu Ala Met Pro Val Glu Tyr Asn Phe Ile Trp Val His 450 455 460
His Pro Trp Ala Phe Gly Asp Ala Gly Cys Arg Gly Tyr Tyr Phe Leu 465 470 475 480
Arg Asp Ala Cys Thr Tyr Ala Thr Ala Leu Asn Val Ala Ser Leu Ser 485 490 495
Val Glu Arg Tyr Leu Ala Ile Cys His Pro Phe Lys Ala Lys Thr Leu 500 505 510
Met Ser Arg Ser Arg Thr Lys Phe Ile Ser Ala Ile Trp Leu Ala 515 520 525
Ser Ala Leu Leu Ala Ile Pro Met Leu Phe Thr Met Gly Leu Gln Asn 530 535 540
Arg Ser Gly Asp Gly Thr His Pro Gly Leu Val Cys Thr Pro Ile 545 550 555 560
Val Asp Thr Ala Thr Val Lys Val Val Ile Gin Val Asn Thr Phe Met 565 570 575
Ser Phe Leu Phe Pro Met Leu Val Ile Ser Ile Leu Asn Thr Val Ile 580 585 590
Ala Asn Lys Leu Thr Val Met Val His Gin Ala Ala Gln Gin Gly Arg 595 600 605
Val Cys Thr Val Gly Thr His Asn Gin Leu Leu His Ser Thr Phe Ann 610 615 620
Met Thr Ile Glu Pro Gly Arg Val Gin Ala Leu Arg His Gly Val Leu 625 630 635 640
Val Leu Arg Ala Val Ile Ala Phe Val Val Cys Trp Leu Pro Tyr 645 650 655
His Val Arg Arg Leu Met Phe Cys Tyr Ile Ser Asp Glu Gin Trp Thr 660 665 670
Thr Phe Leu Phe Asp Phe Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala 675 680 685
Leu Phe Tyr Val Ser Ser Ala Ile Asn Pro Ile Leu Tyr Asn Leu Val 690 695 700
Ser Ala Asn Phe Arg Gin Val Phe Leu Ser Thr Leu Ala Cys Leu Cys 705 710 715 720
Pro Gly Trp Arg Arg Arg Lys Arg Arg Pro Thr Phe Ser Arg Lys 725 730 735
Pro Asn Ser Met Ser Ser Ser His Ala Phe Ser Thr Ser Ala Thr Arg 740 745 750
Glu Thr Leu Tyr Ala Ala Ala Asp Tyr Lys Asp Asp Asp Asp Lys 755 760 765

<210> SEQ ID NO 304
<211> LENGTH: 425
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 304

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
| Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg |
| Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn Ser Asp Leu Asp Val Asn |
| Thr Asp Ile Tyr Ser Lys Val Leu Val Thr Ala Ile Tyr Leu Ala Leu |
| Phe Val Val Gly Thr Val Gly Asn Ser Val Thr Ala Phe Thr Leu Ala |
| Arg Lys Lys Ser Leu Gln Ser Leu Gln Ser Thr Val His Tyr His Leu |
| Gly Ser Leu Ala Leu Ser Leu Leu Leu Leu Leu Ala Leu Ala Met Pro |
| Val Glu Leu Tyr Asn Phe Ile Trp Val His His Pro Trp Ala Phe Gly |
| Asp Ala Gly Cys Arg Gly Tyr Tyr Phe Leu Arg Asp Ala Cys Thr Tyr |
| Ala Thr Ala Leu Asn Val Ala Ser Leu Ser Val Glu Arg Tyr Leu Ala |
| Ala Cys His Pro Phe Lys Ala Lys Thr Leu Met Ser Arg Ser Arg Thr |
| Ile Cys His Pro Phe Lys Ala Lys Thr Leu Met Ser Arg Ser Arg Thr |
| Lys Lys Phe Ile Ser Ala Ile Trp Leu Ala Ser Ala Leu Leu Ala Ile |
| Pro Met Leu Phe Thr Met Gly Leu Gln Asn Arg Ser Gly Asp Gly Thr |
| His Pro Gly Gly Leu Val Cys Thr Pro Ile Val Asp Thr Ala Thr Val |
| Lys Val Val Ile Gln Val Asn Thr Phe Met Ser Phe Leu Phe Pro Met |
| Leu Val Ile Ser Ile Leu Asn Thr Val Ile Ala Asn Lys Leu Thr Val |
| Met Val His Gln Ala Ala Ala Glu Gln Gly Arg Val Cys Thr Val Gly Thr |
| His Asn Gly Leu Glu His Ser Thr Phe Asn Met Thr Ile Glu Pro Gly |
| Arg Val Gln Ala Leu Arg His Gly Val Leu Arg Ala Val Val |
| Ile Ala Phe Val Val Cys Trp Leu Pro Tyr His Val Arg Arg Leu Met |
| Phe Cys Tyr Ile Ser Asp Glu Gln Trp Thr Thr Phe Leu Phe Asp Phe |
| Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala Leu Phe Tyr Val Ser Ser |
| Ala Ile Asn Pro Ile Leu Tyr Asn Leu Val Ser Ala Asn Phe Arg Gln |
| Val Phe Leu Ser Thr Leu Ala Cys Leu Cys Pro Gly Trp Arg His Arg |
| Arg Lys Lys Arg Pro Thr Phe Ser Arg Lys Pro Asn Ser Met Ser Ser |
| Asn His Ala Phe Ser Thr Ser Ala Thr Arg Glu Thr Leu Tyr Ala Ala |
Ala Asp Tyr Lys Asp Asp Asp Asp Lys
420 425

<210> SEQ ID NO 305
<211> LENGTH: 878
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 305

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1      5      10     15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20     25     30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35     40     45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50     55     60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65     70     75     80
Asp Gly Pro Asp Ile Lys Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85     90     95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100    105    110
Asp Lys Leu Tyr Pro Phe Thr Pro Asp Val Arg Tyr Asn Gly Lys
115    120    125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130    135    140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145    150    155    160
Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
165    170    175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180    185    190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195    200    205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210    215    220
Ile Lys Asn His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225    230    235    240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245    250    255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260    265    270
Leu Pro Thr Phe Lys Gly Gin Pro Ser Lys Pro Phe Val Gly Val Leu
275    280    285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290    295    300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Ala Val Asn
305    310    315    320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325    330    335
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys</td>
<td>340 345 350</td>
</tr>
<tr>
<td>Gly Glu Ile Met Pro Asn Ile Pro Glu Met Ser Ala Phe Trp Tyr Ala</td>
<td>355 360 365</td>
</tr>
<tr>
<td>Val Leu Ile Glu Ala Arg Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn</td>
<td>370 375 380</td>
</tr>
<tr>
<td>Ser Asp Leu Asp Val Asn Thr Asp Ile Tyr Ser Lys Val Leu Val Thr</td>
<td>385 390 395 400</td>
</tr>
<tr>
<td>Ala Ile Tyr Leu Ala Leu Phe Val Val Gly Thr Val Gly Asn Ser Val</td>
<td>405 410 415</td>
</tr>
<tr>
<td>Thr Ala Phe Thr Leu Ala Arg Lys Lys Ser Leu Gln Ser Leu Gln Ser</td>
<td>420 425 430</td>
</tr>
<tr>
<td>Thr Val His Tyr His Leu Gly Ser Leu Ala Leu Ser Asp Leu Leu Ile</td>
<td>435 440 445</td>
</tr>
<tr>
<td>Leu Leu Leu Ala Met Pro Val Glu Leu Tyr Asn Phe Ile Trp Val His</td>
<td>450 455 460</td>
</tr>
<tr>
<td>His Pro Trp Ala Phe Gly Asp Ala Gly Cys Arg Gly Tyr Tyr Phe Leu</td>
<td>465 470 475 480</td>
</tr>
<tr>
<td>Arg Asp Ala Cys Thr Tyr Ala Thr Ala Leu Asn Val Ala Ser Leu Ser</td>
<td>485 490 495</td>
</tr>
<tr>
<td>Val Glu Arg Tyr Leu Ala Ile Cys His Pro Phe Lys Ala Lys Thr Leu</td>
<td>500 505 510</td>
</tr>
<tr>
<td>Met Ser Arg Ser Arg Thr Lys Phe Ile Ser Ala Ile Trp Leu Ala</td>
<td>515 520 525</td>
</tr>
<tr>
<td>Ser Ala Leu Leu Ala Ile Pro Met Leu Phe Thr Met Gly Leu Gln Asn</td>
<td>530 535 540</td>
</tr>
<tr>
<td>Arg Ser Gly Asp Gly Thr His Pro Gly Gly Leu Val Cys Thr Pro Ile</td>
<td>545 550 555 560</td>
</tr>
<tr>
<td>Val Asp Thr Ala Thr Val Lys Val Val Ile Gin Val Asn Thr Phe Met</td>
<td>565 570 575</td>
</tr>
<tr>
<td>Ser Phe Leu Phe Pro Met Leu Val Ile Ser Ile Leu Asn Thr Val Ile</td>
<td>580 585 590</td>
</tr>
<tr>
<td>Ala Asn Lys Leu Thr Val Met Val His Gin Ala Ala Gln Gln Gly Arg</td>
<td>595 600 605</td>
</tr>
<tr>
<td>Val Cys Thr Val Gly Thr His Asn Gly Leu Glu His Ser Thr Phe Asn</td>
<td>610 615 620</td>
</tr>
<tr>
<td>Met Thr Ile Glu Pro Gly Arg Val Gin Ala Leu Arg His Gly Val Leu</td>
<td>625 630 635 640</td>
</tr>
<tr>
<td>Val Leu Arg Ala Val Ile Ala Phe Val Val Cys Trp Leu Pro Tyr</td>
<td>645 650 655</td>
</tr>
<tr>
<td>His Val Arg Arg Leu Met Phe Cys Tyr Ile Ser Asp Glu Gin Trp Thr</td>
<td>660 665 670</td>
</tr>
<tr>
<td>Thr Phe Leu Phe Asp Phe Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala</td>
<td>675 680 685</td>
</tr>
<tr>
<td>Leu Phe Tyr Val Ser Ser Ala Asn Pro Ile Leu Tyr Asn Leu Val</td>
<td>690 695 700</td>
</tr>
<tr>
<td>Ser Ala Asn Phe Arg Gln Val Phe Leu Ser Thr Leu Ala Cys Leu Cys</td>
<td>705 710 715 720</td>
</tr>
<tr>
<td>Pro Gly Trp Arg His Arg Arg Lys Lys Arg Pro Thr Phe Ser Arg Lys</td>
<td>725 730 735</td>
</tr>
<tr>
<td>Pro Asn Ser Met Ser Ser Asn His Ala Phe Ser Thr Ser Ala Thr Arg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>740</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>Glu Thr Leu Tyr Ala Ala Ala Ser Asp Lys Ile Ile His Leu Thr Asp</td>
</tr>
<tr>
<td>2</td>
<td>Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val</td>
</tr>
<tr>
<td>3</td>
<td>Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile</td>
</tr>
<tr>
<td>4</td>
<td>Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys</td>
</tr>
<tr>
<td>5</td>
<td>Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg</td>
</tr>
<tr>
<td>6</td>
<td>Gly Ile Pro Thr Leu Leu Lys Phe Lys Asn Gly Glu Val Ala Ala Thr</td>
</tr>
<tr>
<td>7</td>
<td>Lys Val Gly Ala Leu Ser Lys Gly Gly Leu Lys Gly Phe Leu Asp Ala</td>
</tr>
<tr>
<td>8</td>
<td>Asn Leu Ala Ala Ala Ala Asp Tyr Lys Asp Asp Asp Lys</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 306
<211> LENGTH: 836
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<222> FEATURE: OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 306

```
    1  Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
    5  10  15
   20  Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
   25  30
   35  Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn Ser Asp Leu Asp Val Asn
   40  45
   50  Thr Asp Ile Tyr Ser Lys Val Leu Val Thr Ala Ile Tyr Leu Ala Leu
   55  60
   65  Phe Val Val Gly Thr Val Gly Asn Ser Val Thr Ala Phe Thr Leu Ala
   70  75  80
   85  Arg Lys Lys Ser Leu Gln Ser Leu Gln Ser Thr Val His Tyr His Leu
   90  95
  100  Gly Ser Leu Ala Leu Ser Asp Leu Leu Ile Leu Leu Leu Ala Met Pro
  105  110
  115  Val Glu Leu Tyr Asn Phe Ile Trp Val His His Pro Trp Ala Phe Gly
  120  125
  130  Asp Ala Gly Cys Arg Gly Tyr Tyr Phe Leu Arg Asp Ala Cys Thr Tyr
  135  140
  145  Ala Thr Ala Leu Lys Val Ala Lys Ala Ser Leu Val Glu Arg Tyr Leu Ala
  150  155  160
  165  Ile Cys His Pro Phe Lys Ala Lys Thr Leu Met Ser Arg Ser Arg Thr
  170  175
  180  Lys Lys Phe Ile Ser Ala Ile Trp Leu Ala Ser Ala Leu Ala Leu Ala Ile
  185  190
  195  Pro Met Leu Phe Thr Met Gly Leu Gln Asn Arg Ser Gly Asp Gly Thr
  200  205
  210  His Pro Gly Gly Leu Val Cys Thr Pro Ile Val Asp Thr Ala Thr Val
```
<table>
<thead>
<tr>
<th></th>
<th>210</th>
<th>215</th>
<th>220</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys Val Val Ile Gln Val Asn Thr Phe Met Ser Phe Leu Phe Pro Met</td>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Leu Val Ile Ser Ile Leu Asn Thr Val Ile Ala Asn Lys Leu Thr Val</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Met Val His Gln Ala Ala Glu Gln Gly Arg Val Cys Thr Val Gly Thr</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>His Asn Gly Leu Glu His Ser Thr Phe Asn Met Thr Ile Glu Pro Gly</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Arg Val Gln Ala Leu Arg His Gly Val Leu Val Leu Arg Ala Val Val</td>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Ile Ala Phe Val Val Cys Trp Leu Pro Tyr His Val Arg Arg Leu Met</td>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Phe Cys Tyr Ile Ser Asp Glu Gln Trp Thr Thr Phe Leu Phe Asp Phe</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala Leu Phe Tyr Val Ser Ser</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Ala Ile Asn Pro Ile Leu Tyr Asn Leu Val Ser Ala Asn Phe Arg Gln</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Val Phe Leu Ser Thr Leu Ala Cys Leu Cys Pro Gly Trp Arg His Arg</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Arg Lys Lys Arg Pro Thr Phe Ser Arg Lys Pro Asn Ser Met Ser Ser</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Asn His Ala Phe Ser Thr Ser Ala Thr Arg Glu Thr Leu Tyr Ala Ala</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Ala Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Thr Ala Glu Trp</td>
<td>430</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Glu Tyr Cln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn</td>
<td>465</td>
<td>470</td>
<td>475</td>
</tr>
<tr>
<td>Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu</td>
<td>485</td>
<td>490</td>
<td>495</td>
</tr>
<tr>
<td>Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser</td>
<td>500</td>
<td>505</td>
<td>510</td>
</tr>
<tr>
<td>Lys Gly Glu Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Ala Ala Ala</td>
<td>515</td>
<td>520</td>
<td>525</td>
</tr>
<tr>
<td>Asp Tyr Lys Aasp Aasp Aasp Aasp Lys</td>
<td>530</td>
<td>535</td>
<td>535</td>
</tr>
</tbody>
</table>

-continued

**SEQ ID NO:** 307
**LENGTH:** 380
**TYPE:** PRT
**ORGANISM:** Artificial Sequence
**FEATURE:** Synthetic amino acid sequence
**SEQUENCE:** 307

```
Met Gly Cys Leu Gly Asn Ser Lys Thr Glu Asp Gln Arg Asn Glu Glu
1  5  10  15
Lys Ala Gln Arg Glu Ala Asn Lys Lys Ile Glu Lys Gln Leu Gln Lys
```
Asp Lys Gln Val Tyr Arg Ala Thr His Arg Leu Leu Leu Leu Gly Ala
35 40 45
Gly Glu Ser Gly Lys Ser Thr Ile Val Lys Gin Met Arg Ile Leu His
50 55 60
Val Asn Gly Phe Asn Gly Asp Ser Glu Lys Ala Thr Lys Val Gin Asp
65 70 75 80
Ile Lys Asn Asn Leu Lys Glu Ala Ile Glu Thr Ile Val Ala Ala Met
95 99 95
Ser Asn Leu Val Pro Pro Val Glu Leu Ala Asn Pro Glu Asn Gin Phe
100 105 110
Arg Val Asp Tyr Ile Leu Ser Val Met Asn Val Pro Asp Phe Asp Phe
115 120 125
Pro Pro Glu Phe Tyr Glu His Ala Lys Ala Leu Thr Glu Asp Glu Gly
130 135 140
Val Arg Ala Cys Tyr Glu Arg Ser Asn Glu Tyr Gin Leu Ile Asp Cys
145 150 155 160
Ala Gin Tyr Phe Leu Asp Lys Ile Asp Val Ile Lys Gin Ala Asp Tyr
165 170 175
Val Pro Ser Asp Gin Asp Leu Ala Arg Cys Arg Val Leu Thr Ser Gly
180 185 190
Ile Phe Glu Thr Lys Phe Gin Val Asp Lys Val Asn Phe His Met Phe
195 200 205
Asp Val Gly Gin Arg Asp Gin Arg Gin Lys Trp Ile Gin Cys Phe
210 215 220
Asn Asp Val Thr Ala Ile Ile Phe Val Val Ala Ser Ser Ser Tyr Asn
225 230 235 240
Met Val Ile Arg Glu Asp Asn Gin Thr Asn Arg Leu Gin Glu Ala Leu
245 250 255
Asn Leu Phe Lys Ser Ile Trp Asn Asn Arg Thr Leu Arg Thr Ile Ser
260 265 270
Val Ile Leu Phe Leu Asn Lys Gin Asp Leu Ala Glu Lys Val Leu
275 280 285
Ala Gly Lys Ser Lys Ile Glu Asp Tyr Phe Pro Glu Phe Ala Arg Tyr
290 295 300
Thr Thr Pro Glu Asp Ala Thr Pro Glu Pro Gly Glu Asp Pro Arg Val
305 310 315 320
Thr Arg Ala Lys Tyr Phe Ile Arg Asp Glu Phe Leu Arg Ile Ser Thr
325 330 335
Ala Ser Gly Asp Gly Arg His Tyr Cys Tyr Pro His Phe Thr Cys Ala
340 345 350
Val Asp Thr Glu Asn Ile Arg Arg Val Phe Asn Arg Arg Asp Ile
355 360 365
Ile Gin Arg Met His Leu Arg Gin Tyr Glu Leu Leu
370 375 380

<210> SEQ ID NO 300
<211> LENGTH: 394
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
-continued

<210> SEQ ID NO 309
<211> LENGTH: 359
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 309

Met Thr Leu Glu Ser Ile Met Ala Cys Cys Leu Ser Glu Glu Ala Lys 1 5 10 15
  Glu Ala Arg Arg Ile Asn Asp Glu Ile Glu Arg Gin Leu Arg Arg Asp 20 25 30
  Lys Arg Asp Ala Arg Arg Glu Leu Leu Leu Leu Gly Thr Gly 35 40 45
  Glu Ser Gly Lys Ser Thr Phe Ile Lys Gin Met Arg Ile Ile His Gly 50 55 60
  Ser Gly Tyr Ser Asp Glu Asp Lys Arg Gly Phe Thr Lys Leu Val Tyr 65 70 75 80
  Gln Asn Ile Phe Thr Ala Met Gin Ala Met Ile Arg Ala Met Asp Thr 85 90 95
  Leu Lys Ile Pro Tyr Lys Tyr Glu His Asn Lys Ala His Ala Gin Leu 100 105 110
  Val Arg Glu Val Asp Val Glu Lys Val Ser Ala Phe Gin Asn Pro Tyr 115 120 125
  Val Asp Ala Ile Lys Ser Leu Trp Asn Asp Pro Gly Ile Gin Glu Cys 130 135 140
  Tyr Asp Arg Arg Arg Glu Tyr Gin Leu Ser Asp Ser Thr Lys Tyr Tyr 145 150 155 160
  Leu Asn Asp Leu Asp Arg Val Ala Asp Pro Ala Tyr Leu Pro Thr Gin 165 170 175
  Gln Asp Val Leu Arg Val Arg Val Arg Val Arg Met Gin Val Asp Arg 180 185 190
  Pro Phe Asp Leu Gin Ser Val Ile Phe Arg Met Val Gin Asp Val Gly Gly 195 200 205
  Gin Arg Ser Glu Arg Arg Lys Trp Asp His Cys Phe Gin Asn Val Thr 210 215 220
  Ser Ile Met Phe Leu Val Ala Leu Ser Glu Tyr Asp Gin Val Leu Val 225 230 235 240
  Glu Ser Asp Gin Glu Arg Met Glu Glu Ser Lys Ala Leu Phe Arg 245 250 255
  Thr Ile Ile Thr Tyr Pro Trp Phe Gin Asn Ser Ser Val Ile Leu Phe 260 265 270
  Leu Asn Lys Lys Asp Leu Leu Glu Glu Lys Ile Met Tyr Ser His Leu 275 280 285
  Val Asp Tyr Phe Pro Glu Tyr Asp Gly Pro Gin Arg Asp Ala Gin Ala 290 295 300
  Ala Arg Glu Phe Ile Leu Lys Met Phe Val Asp Leu Asn Pro Asp Ser 305 310 315 320
  Asp Lys Ile Ile Tyr Ser His Phe Thr Cys Ala Thr Asp Thr Glu Asn 325 330 335
  Ile Arg Phe Val Phe Ala Ala Val Lys Asp Thr Ile Leu Gin Leu Asn 340 345 350

May 31, 2012
Leu Lys Glu Tyr Asn Leu Val

<210> SEQ ID NO 310
<211> LENGTH: 355
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 310

Met Gly Cys Thr Val Ser Ala Glu Asp Lys Ala Ala Ala Glu Arg Ser
1      5      10     15
Lys Met Ile Asp Lys Ala Arg Glu Asp Gly Glu Ala Ala Arg
20     25     30
Glu Val Lys Leu Leu Leu Gly Ala Gly Ser Gly Lys Ser Thr
35     40     45
Ile Val Lys Glu Met Lys Ile Ile His Glu Asp Gly Tyr Ser Glu Glu
50     55     60
Glu Cys Arg Gln Tyr Arg Ala Val Val Tyr Ser Asn Thr Ile Gln Ser
65     70     75     80
Ile Met Ala Ile Val Lys Ala Met Gly Asn Leu Gln Ile Asp Phe Ala
85     90     95
Asp Pro Ser Arg Ala Asp Ala Arg Gln Leu Phe Ala Leu Ser Cys
100    105    110
Thr Ala Glu Glu Gln Gly Val Leu Pro Asp Asp Leu Ser Gly Val Ile
115    120    125
Arg Arg Leu Trp Ala Asp His Gly Val Gin Ala Cys Phe Gly Arg Ser
130    135    140
Arg Glu Tyr Gln Leu Asn Asp Ser Ala Ala Tyr Tyr Leu Asn Asp Leu
145    150    155    160
Glu Arg Ile Ala Gln Ser Asp Tyr Ile Pro Thr Gin Gln Asp Val Leu
165    170    175
Arg Thr Arg Val Lys Thr Thr Gly Ile Val Glu Thr His Phe Thr Phe
180    185    190
Lys Asp Leu His Phe Lys Met Phe Asp Val Gly Gly Gin Arg Ser Glu
195    200    205
Arg Lys Lys Trp Ile His Cys Phe Glu Gly Val Thr Ala Ile Ile Phe
210    215    220
Cys Val Ala Leu Ser Ala Tyr Asp Leu Val Leu Ala Glu Asp Glu Val
225    230    235    240
Met Asn Arg Met His Glu Ser Met Lys Leu Phe Asp Ser Ile Cys Asn
245    250   255
Asn Lys Trp Phe Thr Asp Thr Ser Ile Leu Phe Leu Asn Lys Lys
260    265    270
Asp Leu Phe Glu Glu Lys Ile Thr His Ser Pro Leu Thr Ile Cys Phe
275    280    285
Pro Glu Tyr Thr Gly Ala Asn Tyr Asp Glu Ala Ala Ser Tyr Ile
290    295    300
Gln Ser Lys Phe Glu Asp Leu Asn Lys Arg Lys Asp Thr Lys Glu Ile
305    310    315    320
Tyr Thr His Phe Thr Cys Ala Thr Asp Thr Lys Asn Val Glu Phe Val
325    330    335
Phe Asp Ala Val Thr Asp Val Ile Lys Asn Asn Leu Lys Asp Cys
340  345  350
Gly Leu Phe
355

<210> SEQ ID NO 311
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 311
Met Ser Gly Val Val Arg Thr Ser Arg Cys Leu Leu Pro Ala Glu
1       5       10      15
Ala Gly Gly Ala Arg Glu Arg Ala Gly Ser Gly Ala Arg Asp Ala
20      25      30
Glu Arg Glu Ala Arg Arg Ser Arg Asp Ile Asp Ala Leu Leu Ala
35      40      45
Arg Glu Arg Arg Ala Val Arg Leu Val Lys Ile Leu Leu Leu Gly
50      55      60
Ala Gly Glu Ser Gly Ser Lys Ser Thr Phe Leu Leu Gly Gln Met Arg Ile Ile
65      70      75      80
His Gly Arg Glu Phe Asp Glu Lys Ala Leu Leu Glu Phe Arg Asp Thr
85      90      95
Ile Phe Asp Asn Ile Leu Lys Gly Ser Arg Val Leu Val Asp Ala Arg
100     105     110
Asp Lys Leu Gly Ile Pro Trp Glu Tyr Ser Glu Asn Glu Lys His Gly
115     120     125
Met Phe Leu Met Ala Phe Glu Asn Lys Ala Gly Leu Pro Val Glu Pro
130     135     140
Ala Thr Phe Gln Leu Tyr Val Pro Ala Leu Ser Ala Leu Thr Arg Asp
145     150     155     160
Ser Gly Ile Arg Glu Ala Phe Ser Arg Ser Glu Phe Gln Leu Gly
165     170     175
Glu Ser Val Lys Tyr Phe Leu Asp Asn Leu Asp Arg Ile Gly Gln Leu
180     185     190
Asn Tyr Phe Pro Ser Lys Gln Asp Ile Leu Leu Ala Arg Lys Ala Thr
195     200     205
Lys Gly Ile Val Glu His Asp Phe Val Ile Lys Lys Ile Pro Phe Lys
210     215     220
Met Val Asp Val Gly Gln Arg Ser Gln Arg Glu Tyr Lys Trp Phe Gln
225     230     235     240
Cys Phe Asp Gly Ile Thr Ser Ile Leu Phe Met Val Ser Ser Ser Glu
245     250     255
Tyr Asp Gln Val Leu Met Glu Asp Arg Arg Thr Asn Arg Leu Val Glu
260     265     270
Ser Met Asn Ile Phe Glu Thr Ile Val Asn Asn Lys Leu Phe Phe Asn
275     280     285
Val Ser Ile Ile Leu Phe Leu Asn Met Asp Leu Leu Val Glu Lys
290     295     300
Val Lys Thr Val Ser Ile Lys Lys His Phe Pro Asp Phe Arg Gly Asp
305     310     315     320
Pro His Gln Leu Glu Asp Val Gln Arg Tyr Leu Val Gln Cys Phe Asp
325 330 335
Arg Lys Arg Arg Asn Arg Ser Lys Pro Leu Phe His His Phe Thr Thr
340 345 350
Ala Ile Asp Thr Glu Asn Val Arg Phe Val Phe His Ala Val Lys Asp
355 360 365
Thr Ile Leu Gln Glu Asn Leu Lys Asp Ile Met Leu Gln
370 375 380

<210> SEQ ID NO 312
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 312
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5 10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20  25  30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp
35  40  45
Asp Lys
50

<210> SEQ ID NO 313
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 313
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys
20  25

<210> SEQ ID NO 314
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 314
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5 10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20  25  30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp
35  40  45
Asp Lys
50

<210> SEQ ID NO 315
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 315

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
      1      5     10     15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
      20    25     30
Leu Gin Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp
      35    40     45
Asp Lys
      50

<210> SEQ ID NO 316
<211> LENGTH: 343
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 316

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
      1      5     10     15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Gly Lys
      20    25     30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
      35    40     45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
      50    55     60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Glu Val Ala Ala Thr Gly
       65     70     75     80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
       85    90     95
Ala Gin Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
      100    105     110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Val Arg Tyr Asn Gly Lys
      115   120    125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
      130   135    140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
      145   150    155    160
Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
      165    170    175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
      180    185     190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
      195   200    205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
      210   215    220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
      225   230    235    240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
      245   250    255
-continued

Ala Trp Ser Arg Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val  
260  265  270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu  
275  280  285
Ser Ala Gly Ile Asn Ala Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu  
290  295  300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn  
305  310  315  320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu  
325  330  335
Leu Ala Lys Arg Pro Arg Ile  
340

<210> SEQ ID NO 317
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 317

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr  
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys  
20  25  30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu  
35  40
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu  
50  55  60
His Pro Asp Lys Leu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly  
65  70  75  80
Asp Gly Pro Asp Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr  
85  90  95
Ala Gin Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin  
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys  
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn  
130 135 140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala  
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Lys Glu Ser Ala Leu Met Phe Asn  
165 170 175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Asp Gly Gly  
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly  
195 200 205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu  
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu  
225 230 235 240
Ala Ala Asn Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp  
245 250 255
### Synthetic Amino Acid Sequence

<table>
<thead>
<tr>
<th>Ala</th>
<th>Trp</th>
<th>Ser</th>
<th>Asn</th>
<th>Ile</th>
<th>Asp</th>
<th>Thr</th>
<th>Ser</th>
<th>Lys</th>
<th>Val</th>
<th>Asn</th>
<th>Tyr</th>
<th>Gly</th>
<th>Val</th>
<th>Thr</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Thr</td>
<td>Phe</td>
<td>Lys</td>
<td>Gly</td>
<td>Gln</td>
<td>Pro</td>
<td>Ser</td>
<td>Lys</td>
<td>Pro</td>
<td>Phe</td>
<td>Val</td>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Gly</td>
<td>Ile</td>
<td>Asn</td>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td>Pro</td>
<td>Asn</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>290</td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Glu</td>
<td>Asn</td>
<td>Tyr</td>
<td>Leu</td>
<td>Thr</td>
<td>Leu</td>
<td>Thr</td>
<td>Asp</td>
<td>Glu</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Lys</td>
<td>Pro</td>
<td>Leu</td>
<td>Gly</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td>Ser</td>
<td>Tyr</td>
<td>Glu</td>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Ile</td>
<td>Ala</td>
<td>Thr</td>
<td>Met</td>
<td>Glu</td>
<td>Asn</td>
<td>Ala</td>
<td>Gln</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Trp</td>
<td>Tyr</td>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Ile</td>
<td>Glu</td>
<td>Ala</td>
<td>Arg</td>
<td>Leu</td>
<td>Gln</td>
<td>Ala</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Glu</td>
<td>Ser</td>
<td>Arg</td>
<td>Asp</td>
<td>Tyr</td>
<td>Lys</td>
<td>Asp</td>
<td>Asp</td>
<td>Asp</td>
<td>Asp</td>
<td>Lys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Other Information:**
- Synthetic amino acid sequence
- Organism: Artificial Sequence
- Feature: Synthetic Sequence
- Sequence: 318

### Synthetic Amino Acid Sequence (Continued)

<table>
<thead>
<tr>
<th>Met</th>
<th>Lys</th>
<th>Ile</th>
<th>Lys</th>
<th>Thr</th>
<th>Gly</th>
<th>Ala</th>
<th>Arg</th>
<th>Ile</th>
<th>Leu</th>
<th>Ala</th>
<th>Leu</th>
<th>Ser</th>
<th>Ala</th>
<th>Leu</th>
<th>Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Met</td>
<td>Met</td>
<td>Phe</td>
<td>Ser</td>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Ile</td>
<td>Gly</td>
<td>Glu</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Ile</td>
<td>Thr</td>
<td>Gly</td>
<td>Ala</td>
<td>Asp</td>
<td>Lys</td>
<td>Gly</td>
<td>Tyr</td>
<td>Asn</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Gly</td>
<td>Lys</td>
<td>Phe</td>
<td>Leu</td>
<td>Gly</td>
<td>Lys</td>
<td>Thr</td>
<td>Gly</td>
<td>Ile</td>
<td>Lys</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Pro</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Glu</td>
<td>Gly</td>
<td>Phe</td>
<td>Pro</td>
<td>Gin</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Thr</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Gly</td>
<td>Pro</td>
<td>Asp</td>
<td>Ile</td>
<td>Phe</td>
<td>Thr</td>
<td>Ala</td>
<td>His</td>
<td>Asp</td>
<td>Arg</td>
<td>Phe</td>
<td>Gly</td>
<td>Gly</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Gln</td>
<td>Ser</td>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Ile</td>
<td>Thr</td>
<td>Pro</td>
<td>Asp</td>
<td>Lys</td>
<td>Ala</td>
<td>Phe</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Tyr</td>
<td>Pro</td>
<td>Phe</td>
<td>Thr</td>
<td>Asp</td>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Tyr</td>
<td>Asn</td>
<td>Gly</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Tyr</td>
<td>Pro</td>
<td>Ile</td>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
<td>Leu</td>
<td>Ser</td>
<td>Leu</td>
<td>Ile</td>
<td>Tyr</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Leu</td>
<td>Leu</td>
<td>Pro</td>
<td>Asn</td>
<td>Pro</td>
<td>Pro</td>
<td>Thr</td>
<td>Thr</td>
<td>Trp</td>
<td>Glu</td>
<td>Ile</td>
<td>Pro</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Lys</td>
<td>Ala</td>
<td>Lys</td>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Met</td>
<td>Phe</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Gln</td>
<td>Glu</td>
<td>Pro</td>
<td>Tyr</td>
<td>Phe</td>
<td>Thr</td>
<td>Trp</td>
<td>Pro</td>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>180</td>
<td>185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Ala</td>
<td>Phe</td>
<td>Lys</td>
<td>Tyr</td>
<td>Glu</td>
<td>Ala</td>
<td>Gly</td>
<td>Lys</td>
<td>Tyr</td>
<td>Asp</td>
<td>Ile</td>
<td>Lys</td>
<td>Asp</td>
<td>Val</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>195</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Asp</td>
<td>Asn</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Leu</td>
<td>Thr</td>
<td>Phe</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1le Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Ala Phe Asn Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Ala Val Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Asp Lys Pro Leu Gly Ala Ala Leu Lys Ser Tyr Glu Glu Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gln Ala Ser Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp Asp Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 319
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220>FEATURE: OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 319

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Gln Ser Gly Leu Leu Ala Gln Ile Thr Pro Asp Lys Ala Phe Gin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Ser
340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Glu Ala Ser Val
355 360 365
Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp Lys
370 375 380

<210> SEQ ID NO 320
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 320

Ala Ser Val Asp Ala Glu Ser Ser Asp Lys Ile Ile His Leu Thr
1 5 10 15
Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
20 25 30
Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro
35 40 45
Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala
50 55 60
Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile
65 70 75 80
Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gln Val Ala Ala
85 90 95
Thr Lys Val Gly Ala Leu Ser Lys Gly Gin Leu Lys Glu Asn Leu Ala
100 105 110
Asp Tyr Lys Asp Asp Asp Lys
115 120

<210> SEQ ID NO 321
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 323

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15

Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
20  25  30

Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro
35  40  45

Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gin Gly Lys Leu Thr Val Ala
50  55  60

Lys Leu Asn Ile Asp Gin Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile
65  70  75  80

Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala
95  90  95

Thr Lys Val Gly Ala Leu Ser Lys Gly Gin Leu Lys Glu Asn Leu Ala
100 105 110

Asp Tyr Lys Asp Asp Asp Asp Lys
115 120

<210> SEQ ID NO 322
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 322

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15

Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
20  25  30

Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro
35  40  45

Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gin Gly Lys Leu Thr Val Ala
50  55  60

Lys Leu Asn Ile Asp Gin Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile
65  70  75  80

Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala
95  90  95

Thr Lys Val Gly Ala Leu Ser Lys Gly Gin Leu Lys Glu Asn Leu Ala
100 105 110

Asp Tyr Lys Asp Asp Asp Asp Lys
115 120

<210> SEQ ID NO 323
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 323

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15
<table>
<thead>
<tr>
<th>AAs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
</tr>
<tr>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td>Thr</td>
<td>Asp</td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>Lys</td>
<td>Ala</td>
</tr>
<tr>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
</tr>
</tbody>
</table>
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35  40  45
Leu Thr Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala
50  55  60
Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile
65  70  75  80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr
85  90  95
Val Ala Lys Leu Asn Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr
100 105 110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val
115 120 125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu Asn
130 135 140
Leu Ala Asp Tyr Lys Asp Asp Asp Asp Lys
145 150

SEQ ID NO 326
LENGTH: 154
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic amino acid sequence

SEQUENCE: 326
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20  25  30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35  40  45
Leu Thr Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala
50  55  60
Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile
65  70  75  80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr
85  90  95
Val Ala Lys Leu Asn Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr
100 105 110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val
115 120 125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu Asn
130 135 140
Leu Ala Asp Tyr Lys Asp Asp Asp Asp Lys
145 150

SEQ ID NO 327
LENGTH: 154
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic amino acid sequence

SEQUENCE: 327
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
-continued

1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20  25  30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35  40  45
Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala
50  55  60
Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile
65  70  75  80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gin Gly Lys Leu Thr
85  90  95
Val Ala Lys Leu Asn Ile Asp Gin Asn Pro Gly Thr Ala Pro Lys Tyr
100  105  110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Asn Gly Glu Val
115  120  125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gin Leu Lys Glu Asn
130  135  140
Leu Ala Asp Tyr Lys Asp Asp Asp Asp Lys
145  150

&lt;210&gt; SEQ ID NO 328
&lt;211&gt; LENGTH: 485
&lt;212&gt; TYPE: PRT
&lt;213&gt; ORGANISM: Artificial Sequence
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: Synthetic amino acid sequence

&lt;400&gt; SEQUENCE: 328

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Glu Gly Lys
20  25  30
Leu Val Ile Trp Ile Asn Gly Asp Tyr Gin Gly Lys Leu Ala Glu
35  40  45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Leu Thr Val Glu
50  55  60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly
65  70  75  80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Lys Tyr
85  90  95
Ala Gin Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
100  105  110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115  120  125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130  135  140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145  150  155  160
Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Asn
165  170  175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Lys
180  185  190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
Val Asp Aen Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Aen Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Aen Ala Ala Ser Pro Aen Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Thr Met Glu Aen Ala Glu Ser
340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Ala Glu Arg Leu Gln Ala Ser Val
355 360 365
Asp Ala Glu Ser Arg Ser Asp Lys Ile His Leu Thr Asp Asp Ser
370 375 380
Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
385 390 395 400
Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
405 410 415
Glu Ile Ala Asp Glu Tyr Glu Gly Leu Thr Val Ala Lys Leu Asn
420 425 430
Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
435 440 445
Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Thr Lys Val
450 455 460
Gly Ala Leu Ser Lys Gly Glu Leu Ala Asp Tyr Lys
465 470 475 480
Asp Asp Asp Asp Lys
485

<210> SEQ ID NO 329
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 329
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30
Leu Val Ile Trp Ile Aen Gly Asp Gly Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
<table>
<thead>
<tr>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td>Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn</td>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td>Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Ile Pro Ala</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td>Leu Asp Lys Glu Leu Lys Gly Lys Ser Ala Leu Met Phe Asn</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly</td>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td>Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td>Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp</td>
<td>245</td>
<td>250</td>
</tr>
<tr>
<td>Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val</td>
<td>260</td>
<td>265</td>
</tr>
<tr>
<td>Leu Pro Thr Phe Lys Gly Gin Pro Ser Lys Pro Phe Val Gly Val Leu</td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Gly Glu</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td>Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gin Ser</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td>Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gin Ala Ser Val</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td>Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe</td>
<td>385</td>
<td>390</td>
</tr>
<tr>
<td>Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Glu Ile Ala Asp Glu Tyr Glu Lys Leu Thr Val Ala Lys Leu Asn</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Ile Asp Gin Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val</td>
<td>450</td>
<td>455</td>
</tr>
</tbody>
</table>
Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu Asn Leu Ala Asp Tyr Lys
465 470 475 480
Asp Asp Asp Asp Lys
485

<210> SEQ ID NO: 330
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 330
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr Met Met Phe Ser Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30
Leu Val Ile Trp Ile Asn Gly Asp Gly Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly
65 70 75 80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95
Ala Glu Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Asn
165 170 175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Pro Asn Ala Ala Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gin Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Asp Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu 325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Thr Met Glu Asn Ala Gln Ser 340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gln Ala Ser Val 355 360 365
Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser 370 375 380
Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe 385 390 395 400
Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp 405 410 415
Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn 420 425 430
Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile 435 440 445
Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val 450 455 460
Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu Asn Leu Ala Asp Tyr Lys 465 470 475 480
Asp Asp Asp Asp Lys 485

<210> SEQ ID NO 331
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 331
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys 20 25 30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu 35 40 45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu 50 55 60
His Pro Asp Lys Leu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly 65 70 75 80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr 85 90 95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln 100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys 115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn 130 135 140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Ile Pro Ala 145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn 165 170 175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
  180
  185
  190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
  195
  200
  205
Val Asp Asn Ala Gly Ala Gly Ala Gly Leu Thr Phe Leu Val Asp Leu
  210
  215
  220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
  225
  230
  235
  240
 Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
  245
  250
  255
 Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
  260
  265
  270
Leu Pro Thr Phe Lys Gly Glu Pro Ser Lys Pro Phe Val Gly Val Leu
  275
  280
  285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
  290
  295
  300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
  305
  310
  315
  320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
  325
  330
  335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Ser
  340
  345
  350
 Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Glu Ala Ser Val
  355
  360
  365
Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser
  370
  375
  380
Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
  385
  390
  395
  400
 Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
  405
  410
  415
Glu Ile Ala Asp Glu Tyr Glu Gly Lys Leu Thr Val Ala Lys Leu Asn
  420
  425
  430
Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
  435
  440
  445
 Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val
  450
  455
  460
Gly Ala Leu Ser Lys Gly Glu Leu Ala Asp Leu Ala Asp Tyr Lys
  465
  470
  475
  480
Asp Asp Asp Lys
  485

<210> SEQ ID NO 332
<211> LENGTH: 767
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 332

     Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1     5     10     15
     Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20    25    30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
55 60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65 70 75 80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
95 90 95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Gly Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Asn
165 170 175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gin Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asm Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Gly Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gin Lys
340 345 350
Gly Glu Ile Met Pro Asn Ile Pro Gin Met Ser Ala Phe Trp Tyr Ala
355 360 365
Val Leu Ile Glu Ala Arg Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn
370 375 380
Ser Asp Leu Asp Val Asn Thr Asp Ile Tyr Ser Lys Val Leu Val Thr
385 390 395 400
Ala Ile Tyr Leu Ala Leu Phe Val Val Gly Thr Val Gly Asn Ser Val
405 410 415
Thr Ala Phe Thr Leu Ala Arg Lys Ser Leu Gin Ser Leu Gin Ser
420 425 430
Thr Val His Tyr His Leu Gly Ser Leu Ala Leu Ser Asp Leu Leu Ile
435 440 445
Leu Leu Leu Ala Met Pro Val Glu Tyr Asn Phe Ile Trp Val His
450 455 460
His Pro Trp Ala Phe Gly Asp Ala Gly Cys Arg Gly Tyr Phe Leu
465 470 475 480
Arg Asp Ala Cys Thr Tyr Ala Thr Ala Leu Asn Val Ala Ser Leu Ser
485 490 495
Val Glu Arg Tyr Leu Ala Ile Cys His Pro Phe Lys Ala Lys Thr Leu
500 505 510
Met Ser Arg Ser Arg Thr Lys Phe Ile Ser Ala Ile Trp Leu Ala
515 520 525
Ser Ala Leu Leu Leu Ala Ile Pro Met Leu Phe Thr Met Gly Leu Gln Asn
530 535 540
Arg Ser Gly Asp Gly Thr His Pro Gly Gly Leu Val Cys Thr Pro Ile
545 550 555 560
Val Asp Thr Ala Thr Val Lys Val Val Ile Gln Val Asn Thr Phe Met
565 570 575
Ser Phe Leu Phe Pro Met Leu Val Ile Ser Ile Leu Asn Thr Val Ile
580 585 590
Ala Asn Lys Leu Thr Val Met Val His Gln Ala Ala Gln Gly Arg
595 600 605
Val Cys Thr Val Gly Thr His Asn Gly Leu Glu His Ser Thr Phe Asn
610 615 620
Met Thr Ile Glu Pro Gly Arg Val Gln Ala Leu Arg His Gly Val Leu
625 630 635 640
Val Leu Arg Ala Val Val Ile Ala Phe Val Val Cys Trp Leu Pro Tyr
645 650 655
His Val Arg Arg Leu Met Phe Cys Tyr Ile Ser Asp Glu Gln Trp Thr
660 665
Thr Phe Leu Phe Asp Phe Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala
670 675 680 685
Leu Phe Tyr Val Ser Ser Ala Ile Asn Pro Ile Leu Tyr Asn Leu Val
690 695 700
Ser Ala Asn Phe Arg Gln Val Phe Leu Ser Thr Leu Ala Cys Leu Cys
705 710 715 720
Pro Gly Trp Arg His Arg Arg Lys Arg Pro Thr Phe Ser Arg Lys
725 730 735
Pro Asn Ser Met Ser Ser Asn His Ala Phe Ser Thr Ser Ala Thr Arg
740 745 750
Glu Thr Leu Tyr Ala Ala Asp Tyr Lys Asp Asp Asp Asp Lys
755 760 765

<210> SEQ ID NO 333
<211> LENGTH: 425
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 333

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr  Met  Met  Phe  Ser  Ala  Ser  Ala  Leu  Ala  Lys  Ile  Ile  Glu  Ala  Arg  
             20       25       30
Thr  Ser  Glu  Ser  Asp  Thr  Ala  Gly  Pro  Asn  Ser  Asp  Leu  Asp  Val  Asn  
             35       40       45
Thr  Asp  Ile  Tyr  Ser  Lys  Val  Leu  Val  Thr  Ala  Ile  Tyr  Leu  Ala  Leu  
             50       55       60
Phe  Val  Val  Gly  Thr  Val  Gly  Asn  Ser  Val  Thr  Ala  Phe  Thr  Leu  Ala  
             65       70       75       80
Arg  Lys  Lys  Ser  Leu  Gln  Ser  Leu  Gln  Ser  Thr  Val  His  Tyr  His  Leu  
             85       90       95
Gly  Ser  Leu  Ala  Leu  Ser  Leu  Leu  Leu  Ala  Met  Pro  
            100      105      110
Val  Glu  Leu  Tyr  Asn  Phe  Ile  Trp  Val  His  His  Pro  Trp  Ala  Phe  Gly  
            115      120      125
Asp  Ala  Gly  Cys  Arg  Gly  Tyr  Tyr  Phe  Leu  Arg  Asp  Ala  Cys  Thr  Tyr  
            130      135      140
Ala  Thr  Ala  Leu  Arg  Val  Ala  Ala  Ser  Val  Ser  Glu  Arg  Tyr  Leu  Ala  
            145      150      155      160
Ile  Cys  His  Pro  Phe  Lys  Ala  Thr  Leu  Met  Ser  Arg  Ser  Arg  Thr  
            165      170      175
Lys  Lys  Phe  Ile  Ser  Ala  Ile  Trp  Leu  Ala  Ser  Ala  Leu  Leu  Ala  Ile  
            180      185      190
Pro  Met  Leu  Phe  Thr  Met  Leu  Gly  Leu  Gln  Asn  Arg  Ser  Gly  Asp  Gly  Thr  
            195      200      205
His  Pro  Gly  Gly  Leu  Val  Cys  Thr  Pro  Ile  Val  Asp  Thr  Ala  Thr  Val  
            210      215      220
Lys  Val  Val  Ile  Gln  Val  Asn  Thr  Phe  Met  Ser  Phe  Leu  Phe  Pro  Met  
            225      230      235      240
Leu  Val  Ile  Ser  Ile  Leu  Asn  Thr  Val  Ile  Ala  Asn  Lys  Leu  Thr  Val  
            245      250      255
Met  Val  His  Gln  Ala  Ala  Gln  Gly  Arg  Val  Cys  Thr  Val  Gly  Thr  
            260      265      270
His  Asn  Gly  Leu  Glu  His  Ser  Thr  Phe  Asn  Met  Thr  Ile  Glu  Pro  Gly  
            275      280      285
Arg  Val  Gln  Ala  Leu  Arg  His  Gly  Val  Leu  Val  Arg  Ala  Val  
            290      295      300
Ile  Ala  Phe  Val  Val  Cys  Trp  Leu  Pro  Tyr  His  Val  Arg  Arg  Leu  Met  
            305      310      315      320
Phe  Cys  Tyr  Ile  Ser  Asp  Glu  Gln  Trp  Thr  Thr  Phe  Leu  Phe  Asp  Phe  
            325      330      335
Tyr  His  Tyr  Phe  Tyr  Met  Leu  Thr  Asn  Ala  Leu  Phe  Tyr  Val  Ser  Ser  
            340      345      350
Ala  Ile  Asn  Pro  Ile  Leu  Tyr  Asn  Leu  Val  Ser  Ala  Asn  Phe  Arg  Gln  
            355      360      365
Val  Phe  Leu  Ser  Thr  Leu  Ala  Cys  Leu  Cys  Pro  Gly  Trp  Arg  His  Arg  
            370      375      380
Arg  Lys  Lys  Arg  Pro  Thr  Phe  Ser  Arg  Lys  Pro  Asn  Ser  Met  Ser  Ser  
            385      390      395      400
Asn  His  Ala  Phe  Ser  Thr  Ser  Ala  Thr  Arg  Glu  Thr  Leu  Tyr  Ala  Ala  
            405      410      415
Ala  Asp  Tyr  Lys  Asp  Asp  Asp  Asp  Lys
<210> SEQ ID NO 334
<211> LENGTH: 878
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 334

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1    5    10    15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20   25
Leu Val Ile Thr Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35   40   45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50   55   60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65   70   75   80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Gly Tyr
85   90   95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100  105  110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115  120  125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130  135  140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145  150  155  160
Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Asn
165  170  175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180  185  190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195  200  205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210  215  220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225  230  235  240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245  250  255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260  265  270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275  280  285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290  295  300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305  310  315  320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325  330  335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gln Lys
<table>
<thead>
<tr>
<th>Gly</th>
<th>Glu</th>
<th>Ile</th>
<th>Met</th>
<th>Pro</th>
<th>Asn</th>
<th>Ile</th>
<th>Pro</th>
<th>Gln</th>
<th>Met</th>
<th>Ser</th>
<th>Ala</th>
<th>Phe</th>
<th>Trp</th>
<th>Tyr</th>
<th>Ala</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>340</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>365</td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Ile</td>
<td>Glu</td>
<td>Ala</td>
<td>Arg</td>
<td>Thr</td>
<td>Ser</td>
<td>Glu</td>
<td>Ser</td>
<td>Asp</td>
<td>Thr</td>
<td>Ala</td>
<td>Gln</td>
<td>Pro</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>380</td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Leu</td>
<td>Asp</td>
<td>Val</td>
<td>Asn</td>
<td>Thr</td>
<td>Arg</td>
<td>Ile</td>
<td>Tyr</td>
<td>Ser</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>395</td>
<td></td>
<td></td>
<td>390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400</td>
</tr>
<tr>
<td>Ala</td>
<td>Ile</td>
<td>Tyr</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Phe</td>
<td>Val</td>
<td>Val</td>
<td>Gln</td>
<td>Thr</td>
<td>Val</td>
<td>Gln</td>
<td>Asn</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>405</td>
<td></td>
<td></td>
<td>410</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>415</td>
</tr>
<tr>
<td>Thr</td>
<td>Ala</td>
<td>Phe</td>
<td>Thr</td>
<td>Leu</td>
<td>Ala</td>
<td>Arg</td>
<td>Lys</td>
<td>Ser</td>
<td>Leu</td>
<td>Gln</td>
<td>Ser</td>
<td>Leu</td>
<td>Gln</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>420</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>430</td>
</tr>
<tr>
<td>Thr</td>
<td>Val</td>
<td>His</td>
<td>Tyr</td>
<td>His</td>
<td>Leu</td>
<td>Gln</td>
<td>Ser</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Ser</td>
<td>Asp</td>
<td>Leu</td>
<td>Leu</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>440</td>
</tr>
<tr>
<td>Leu</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Met</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Leu</td>
<td>Tyr</td>
<td>Asn</td>
<td>Phe</td>
<td>Ile</td>
<td>Trp</td>
<td>Val</td>
<td>His</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>460</td>
</tr>
<tr>
<td>His</td>
<td>Pro</td>
<td>Trp</td>
<td>Ala</td>
<td>Phe</td>
<td>Gly</td>
<td>Asp</td>
<td>Ala</td>
<td>Gly</td>
<td>Cys</td>
<td>Arg</td>
<td>Gly</td>
<td>Tyr</td>
<td>Tyr</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>480</td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Ala</td>
<td>Cys</td>
<td>Thr</td>
<td>Tyr</td>
<td>Ala</td>
<td>Thr</td>
<td>Ala</td>
<td>Leu</td>
<td>Asn</td>
<td>Val</td>
<td>Ala</td>
<td>Ser</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>485</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>495</td>
</tr>
<tr>
<td>Val</td>
<td>Glu</td>
<td>Arg</td>
<td>Tyr</td>
<td>Leu</td>
<td>Ala</td>
<td>Ile</td>
<td>Cys</td>
<td>His</td>
<td>Pro</td>
<td>Phe</td>
<td>Lys</td>
<td>Ala</td>
<td>Lys</td>
<td>Thr</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>510</td>
</tr>
<tr>
<td>Met</td>
<td>Ser</td>
<td>Arg</td>
<td>Ser</td>
<td>Arg</td>
<td>Thr</td>
<td>Lys</td>
<td>Phe</td>
<td>Ile</td>
<td>Ser</td>
<td>Ala</td>
<td>Ile</td>
<td>Trp</td>
<td>Leu</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>525</td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Ile</td>
<td>Pro</td>
<td>Met</td>
<td>Leu</td>
<td>Phe</td>
<td>Thr</td>
<td>Met</td>
<td>Gln</td>
<td>Gly</td>
<td>Ser</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>530</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>540</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Gly</td>
<td>Asp</td>
<td>Gly</td>
<td>Thr</td>
<td>His</td>
<td>Pro</td>
<td>Gly</td>
<td>Gly</td>
<td>Leu</td>
<td>Val</td>
<td>Cys</td>
<td>Thr</td>
<td>Pro</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>545</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>560</td>
</tr>
<tr>
<td>Val</td>
<td>Asp</td>
<td>Thr</td>
<td>Ala</td>
<td>Thr</td>
<td>Val</td>
<td>Lys</td>
<td>Val</td>
<td>Val</td>
<td>Ile</td>
<td>Gln</td>
<td>Val</td>
<td>Asn</td>
<td>Thr</td>
<td>Phe</td>
<td>Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>565</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>575</td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Leu</td>
<td>Phe</td>
<td>Pro</td>
<td>Met</td>
<td>Leu</td>
<td>Val</td>
<td>Ile</td>
<td>Ser</td>
<td>Ile</td>
<td>Leu</td>
<td>Asn</td>
<td>Thr</td>
<td>Val</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>580</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>590</td>
</tr>
<tr>
<td>Ala</td>
<td>Asn</td>
<td>Lys</td>
<td>Leu</td>
<td>Thr</td>
<td>Val</td>
<td>Met</td>
<td>Val</td>
<td>His</td>
<td>Gln</td>
<td>Ala</td>
<td>Ala</td>
<td>Gly</td>
<td>Gln</td>
<td>Gly</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>595</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>605</td>
</tr>
<tr>
<td>Val</td>
<td>Cys</td>
<td>Thr</td>
<td>Val</td>
<td>Gly</td>
<td>Thr</td>
<td>His</td>
<td>Arg</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>His</td>
<td>Ser</td>
<td>Thr</td>
<td>Phe</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>610</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>620</td>
</tr>
<tr>
<td>Met</td>
<td>Thr</td>
<td>Ile</td>
<td>Glu</td>
<td>Pro</td>
<td>Gly</td>
<td>Arg</td>
<td>Val</td>
<td>Gln</td>
<td>Ala</td>
<td>Arg</td>
<td>His</td>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>625</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>640</td>
</tr>
<tr>
<td>Val</td>
<td>Leu</td>
<td>Arg</td>
<td>Ala</td>
<td>Val</td>
<td>Ile</td>
<td>Ala</td>
<td>Phe</td>
<td>Val</td>
<td>Val</td>
<td>Cys</td>
<td>Thr</td>
<td>Leu</td>
<td>Pro</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>645</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>655</td>
</tr>
<tr>
<td>His</td>
<td>Val</td>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
<td>Met</td>
<td>Phe</td>
<td>Cys</td>
<td>Tyr</td>
<td>Ile</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Gln</td>
<td>Trp</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>660</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>670</td>
</tr>
<tr>
<td>Thr</td>
<td>Phe</td>
<td>Leu</td>
<td>Phe</td>
<td>Asp</td>
<td>Phe</td>
<td>Tyr</td>
<td>His</td>
<td>Tyr</td>
<td>Phe</td>
<td>Tyr</td>
<td>Met</td>
<td>Leu</td>
<td>Thr</td>
<td>Asn</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>675</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>685</td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Tyr</td>
<td>Val</td>
<td>Ser</td>
<td>Ser</td>
<td>Ala</td>
<td>Ile</td>
<td>Asn</td>
<td>Pro</td>
<td>Ile</td>
<td>Leu</td>
<td>Tyr</td>
<td>Asn</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>690</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>700</td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Asn</td>
<td>Phe</td>
<td>Arg</td>
<td>Gln</td>
<td>Val</td>
<td>Phe</td>
<td>Leu</td>
<td>Ser</td>
<td>Thr</td>
<td>Ala</td>
<td>Cys</td>
<td>Leu</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>705</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>720</td>
</tr>
<tr>
<td>Pro</td>
<td>Gly</td>
<td>Trp</td>
<td>Arg</td>
<td>His</td>
<td>Arg</td>
<td>Lys</td>
<td>Lys</td>
<td>Arg</td>
<td>Pro</td>
<td>Thr</td>
<td>Phe</td>
<td>Ser</td>
<td>Arg</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>725</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>735</td>
</tr>
<tr>
<td>Pro</td>
<td>Asn</td>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>Arg</td>
<td>His</td>
<td>Ala</td>
<td>Phe</td>
<td>Ser</td>
<td>Thr</td>
<td>Ser</td>
<td>Ala</td>
<td>Thr</td>
<td>Arg</td>
<td></td>
</tr>
</tbody>
</table>
Glu Thr Leu Tyr Ala Ala Ala Ser Asp Lys Ile Ile His Leu Thr Asp 755 760 765
Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val 770 775 780
Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile 785 790 795 800
Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys 805 810 815
Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg 820 825 830
Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr 835 840 845
Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala 850 855 860
Asn Leu Ala Ala Ala Asp Tyr Lys Asp Asp Asp Lys 865 870 875

<210> SEQ ID NO 335
<211> LENGTH: 536
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 335
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg 20 25 30
Thr Ser Glu Ser Asp Thr Ala Gly Pro Asn Ser Asp Leu Asp Val Asn 35 40 45
Thr Asp Ile Tyr Ser Lys Val Leu Val Thr Ala Ile Tyr Leu Ala Leu 50 55 60
Phe Val Val Gly Thr Val Gly Asn Ser Val Thr Ala Phe Thr Leu Ala 65 70 75 80
Arg Lys Lys Ser Leu Gln Ser Leu Gln Ser Thr Val His Tyr His Leu 85 90 95
Gly Ser Leu Ala Leu Ser Asp Leu Leu Leu Leu Leu Leu Ala Met Pro 100 105 110
Val Glu Leu Tyr Asn Phe Ile Trp Val His His Pro Trp Ala Phe Gly 115 120 125
Asp Ala Gly Cys Arg Gly Tyr Phe Leu Arg Asp Ala Cys Thr Tyr 130 135 140
 Ala Thr Ala Leu Asn Val Ala Ser Leu Ser Val Glu Arg Tyr Leu Ala 145 150 155 160
Ile Cys His Pro Phe Lys Ala Lys Thr Leu Met Ser Arg Ser Arg Thr 165 170 175
Lys Lys Phe Ile Ser Ala Ile Trp Leu Ala Ser Ala Leu Leu Ala Ile 180 185 190
Pro Met Leu Phe Thr Met Gly Leu Glu Asn Arg Ser Gly Asp Gly Thr 195 200 205
His Pro Gly Gly Leu Val Cys Thr Pro Ile Val Asp Thr Ala Thr Val 210 215 220
Lys Val Val Ile Gln Val Asn Thr Phe Met Ser Phe Leu Phe Pro Met
225 230 235 240
Leu Val Ile Ser Ile Leu Asn Thr Val Ile Ala Asn Lys Leu Thr Val
245 250 255
Met Val His Gln Ala Ala Glu Gln Gly Arg Val Cys Thr Val Gly Thr
260 265 270
His Asn Gly Leu Glu His Ser Thr Phe Asn Met Thr Ile Glu Pro Gly
275 280 285
Arg Val Gln Ala Leu Arg His Gly Leu Val Leu Arg Ala Val Val
290 295 300
Ile Ala Phe Val Val Cys Trp Leu Pro Tyr His Val Arg Arg Leu Met
305 310 315 320
Phe Cys Tyr Ile Ser Asp Glu Gln Trp Thr Thr Phe Leu Phe Asp Phe
325 330 335
Tyr His Tyr Phe Tyr Met Leu Thr Asn Ala Leu Phe Tyr Val Ser Ser
340 345 350
Ala Ile Asn Pro Ile Leu Tyr Asn Leu Val Ser Ala Asn Phe Arg Gln
355 360 365
Val Phe Leu Ser Thr Leu Ala Cys Leu Cys Pro Gly Trp Arg His Arg
370 375 380
Arg Lys Lys Arg Pro Thr Phe Ser Arg Lys Pro Asn Ser Met Ser Ser
385 390 395 400
Asn His Ala Phe Ser Thr Ser Ala Thr Arg Glu Thr Leu Tyr Ala Ala
405 410 415
Ala Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp
420 425 430
Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe Trp Ala Glu Trp
435 440 445
Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp
450 455 460
Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp Gln Asn
465 470 475 480
Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu
485 490 495
Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser
500 505 510
Lys Gly Gln Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Ala Ala Ala
515 520 525
Asp Tyr Lys Asp Asp Asp Asp Lys
530 535

&lt;210&gt; SEQ ID NO 336
&lt;211&gt; LENGTH: 380
&lt;212&gt; TYPE: PRT
&lt;213&gt; ORGANISM: Artificial Sequence
&lt;220&gt; FEATURE:
&lt;223&gt; OTHER INFORMATION: Synthetic amino acid sequence
&lt;400&gt; SEQUENCE: 336
Met Gly Cys Leu Gly Asn Ser Lys Thr Glu Asp Gln Arg Asn Glu Glu
1 5 10 15
Lys Ala Gln Arg Glu Ala Asn Lys Lys Ile Glu Lys Gln Leu Gln Lys
20 25 30
Amp Lys Gln Val Tyr Arg Ala Thr His Arg Leu Leu Leu Leu Gly Ala
35  40  45
Gly Glu Ser Gly Lys Ser Thr Ile Val Lys Glu Met Arg Ile Leu His
50  55  60
Val Asn Gly Phe Asn Gly Asp Ser Glu Lys Ala Thr Lys Val Glu Asp
75  80
Ile Lys Asn Asn Leu Lys Glu Ala Ile Glu Thr Ile Val Ala Ala Met
85  90  95
Ser Asn Leu Val Pro Pro Val Glu Leu Ala Asn Pro Glu Asn Glu Phe
100 105 110
Arg Val Asp Tyr Ile Leu Ser Val Met Asn Val Pro Asp Phe Asp Phe
115 120 125
Pro Pro Glu Phe Tyr Glu His Ala Asn Ala Leu Trp Glu Asp Glu Gly
130 135 140
Val Arg Ala Cys Tyr Glu Arg Ser Asn Glu Tyr Glu Leu Ile Asp Cys
145 150 155 160
 Ala Gln Tyr Phe Leu Asp Lys Ile Asp Val Ile Lys Glu Ala Asp Tyr
165 170 175
Val Pro Ser Asp Gln Asp Leu Arg Cys Arg Val Leu Thr Ser Gly
180 185 190
Ile Phe Glu Thr Lys Phe Glu Val Asp Lys Val Asn Phe His Met Phe
195 200 205
Asp Val Gly Gly Glu Arg Arg Glu Arg Lys Trp Ile Glu Cys Phe
210 215 220
Asn Asp Val Thr Ala Ile Ile Phe Val Val Asn Ser Ser Ser Tyr Asn
225 230 235 240
Met Val Ile Arg Glu Asp Asn Glu Thr Asn Arg Leu Glu Gln Ala Leu
245 250 255
Asn Leu Phe Lys Ser Ile Trp Asn Asn Arg Trp Leu Arg Thr Ile Ser
260 265 270
Val Ile Leu Phe Leu Asn Lys Gln Asp Leu Ala Glu Lys Val Leu
275 280 285
 Ala Gly Lys Ser Lys Ile Glu Asp Tyr Phe Pro Glu Phe Ala Arg Tyr
290 295 300
Thr Thr Pro Glu Asp Ala Thr Pro Glu Pro Gly Glu Asp Pro Arg Val
305 310 315 320
Thr Arg Ala Lys Tyr Phe Ile Arg Asp Glu Phe Leu Arg Ile Ser Thr
325 330 335
Ala Ser Gly Asp Gly Arg His Tyr Cys Tyr Pro His Phe Thr Cys Ala
340 345 350
Val Asp Thr Glu Asn Ile Arg Arg Val Phe Asn Asp Cys Arg Asp Ile
355 360 365
Ile Gln Arg Met His Leu Arg Glu Tyr Glu Leu
370 375 380

<210> SEQ ID NO 337
<211> LENGTH: 394
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FRATURR:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 337
<table>
<thead>
<tr>
<th>Met</th>
<th>Gly</th>
<th>Cys</th>
<th>Leu</th>
<th>Gly</th>
<th>Asn</th>
<th>Ser</th>
<th>Lys</th>
<th>Thr</th>
<th>Glu</th>
<th>Asp</th>
<th>Gln</th>
<th>Arg</th>
<th>Asn</th>
<th>Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ala</td>
<td>Gln</td>
<td>Arg</td>
<td>Glu</td>
<td>Ala</td>
<td>Asn</td>
<td>Ser</td>
<td>Lys</td>
<td>1</td>
<td>Glu</td>
<td>Lys</td>
<td>Gln</td>
<td>Leu</td>
<td>Gln</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Lys</td>
<td>Gln</td>
<td>Val</td>
<td>Tyr</td>
<td>Arg</td>
<td>Ala</td>
<td>Thr</td>
<td>His</td>
<td>Arg</td>
<td>Leu</td>
<td>Leu</td>
<td>Leu</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Ser</td>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
<td>Thr</td>
<td>Ile</td>
<td>Val</td>
<td>Lys</td>
<td>Gln</td>
<td>Met</td>
<td>Arg</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Asn</td>
<td>Gly</td>
<td>Phe</td>
<td>Asn</td>
<td>Gly</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Asp</td>
<td>Pro</td>
<td>Glu</td>
<td>Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Asn</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Glu</td>
<td>Ala</td>
<td>Thr</td>
<td>Lys</td>
<td>Val</td>
<td>Gln</td>
<td>Asp</td>
<td>Ile</td>
<td>Lys</td>
</tr>
<tr>
<td>95</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Leu</td>
<td>Lys</td>
<td>Ala</td>
<td>Ile</td>
<td>Glu</td>
<td>Thr</td>
<td>Ile</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Met</td>
<td>Ser</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Val</td>
<td>Pro</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
<td>Asn</td>
<td>Pro</td>
<td>Glu</td>
<td>Asn</td>
<td>Glu</td>
<td>Phe</td>
<td>Arg</td>
<td>Val</td>
</tr>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Tyr</td>
<td>Ile</td>
<td>Leu</td>
<td>Ser</td>
<td>Val</td>
<td>Met</td>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Pro</td>
<td>Asp</td>
<td>Phe</td>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Phe</td>
<td>Tyr</td>
<td>Glu</td>
<td>His</td>
<td>Ala</td>
<td>Lys</td>
<td>Ala</td>
<td>Leu</td>
<td>Trp</td>
<td>Glu</td>
<td>Asp</td>
<td>Glu</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Cys</td>
<td>Tyr</td>
<td>Glu</td>
<td>Arg</td>
<td>Ser</td>
<td>Asn</td>
<td>Glu</td>
<td>Tyr</td>
<td>Gln</td>
<td>Leu</td>
<td>Ile</td>
<td>Asp</td>
<td>Cys</td>
<td>Ala</td>
</tr>
<tr>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Phe</td>
<td>Leu</td>
<td>Asp</td>
<td>Lys</td>
<td>Ile</td>
<td>Asp</td>
<td>Val</td>
<td>Ile</td>
<td>Lys</td>
<td>Gln</td>
<td>Ala</td>
<td>Asp</td>
<td>Tyr</td>
<td>Val</td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Asp</td>
<td>Gln</td>
<td>Asp</td>
<td>Leu</td>
<td>Leu</td>
<td>Arg</td>
<td>Cys</td>
<td>Arg</td>
<td>Val</td>
<td>Leu</td>
<td>Thr</td>
<td>Ser</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Thr</td>
<td>Lys</td>
<td>Phe</td>
<td>Gln</td>
<td>Val</td>
<td>Asp</td>
<td>Lys</td>
<td>Val</td>
<td>Asn</td>
<td>Phe</td>
<td>His</td>
<td>Met</td>
<td>Phe</td>
<td>Asp</td>
</tr>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Gly</td>
<td>Gln</td>
<td>Arg</td>
<td>Asp</td>
<td>Glu</td>
<td>Arg</td>
<td>Lys</td>
<td>Trp</td>
<td>Ile</td>
<td>Gln</td>
<td>Cys</td>
<td>Phe</td>
<td>Asn</td>
<td>Aas</td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Thr</td>
<td>Ala</td>
<td>Ile</td>
<td>Phe</td>
<td>Val</td>
<td>Ala</td>
<td>Ser</td>
<td>Ser</td>
<td>Ser</td>
<td>Tyr</td>
<td>Asn</td>
<td>Met</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Arg</td>
<td>Glu</td>
<td>Asp</td>
<td>Gln</td>
<td>Thr</td>
<td>Asn</td>
<td>Arg</td>
<td>Leu</td>
<td>Gln</td>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
<td>Asn</td>
<td>Leu</td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Lys</td>
<td>Ser</td>
<td>Ile</td>
<td>Trp</td>
<td>Asn</td>
<td>Arg</td>
<td>Trp</td>
<td>Leu</td>
<td>Arg</td>
<td>Thr</td>
<td>Ile</td>
<td>Ser</td>
<td>Val</td>
<td>Ile</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Asn</td>
<td>Lys</td>
<td>Gln</td>
<td>Arg</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>290</td>
<td>285</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ser</td>
<td>Lys</td>
<td>Ile</td>
<td>Glu</td>
<td>Asp</td>
<td>Tyr</td>
<td>Phe</td>
<td>Pro</td>
<td>Glu</td>
<td>Phe</td>
<td>Ala</td>
<td>Arg</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Asp</td>
<td>Ala</td>
<td>Thr</td>
<td>Pro</td>
<td>Glu</td>
<td>Pro</td>
<td>Glu</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Val</td>
<td>Thr</td>
<td>Arg</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Tyr</td>
<td>Phe</td>
<td>Ile</td>
<td>Arg</td>
<td>Asp</td>
<td>Glu</td>
<td>Phe</td>
<td>Leu</td>
<td>Arg</td>
<td>Ile</td>
<td>Ser</td>
<td>Thr</td>
<td>Ala</td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Asp</td>
<td>Gln</td>
<td>Arg</td>
<td>His</td>
<td>Tyr</td>
<td>Cys</td>
<td>Tyr</td>
<td>Pro</td>
<td>His</td>
<td>Phe</td>
<td>Thr</td>
<td>Cys</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>365</td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Glu</td>
<td>Asn</td>
<td>Ile</td>
<td>Arg</td>
<td>Val</td>
<td>Phe</td>
<td>Asn</td>
<td>Arg</td>
<td>Asp</td>
<td>Ile</td>
<td>Ile</td>
<td>Gln</td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Met</td>
<td>His</td>
<td>Leu</td>
<td>Arg</td>
<td>Gln</td>
<td>Tyr</td>
<td>Glu</td>
<td>Leu</td>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
<210> SEQ ID NO 338
<211> LENGTH: 359
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 338

Met Thr Leu Glu Ser Ile Met Ala Cys Cys Leu Ser Glu Glu Ala Lys
1     5     10   15

Glu Ala Arg Arg Ile Asp Glu Ile Glu Arg Glu Leu Arg Arg Asp
20    25    30

Lys Arg Asp Ala Arg Glu Leu Lys Leu Leu Leu Gly Thr Gly
35    40    45

Glu Ser Gly Lys Ser Thr Phe Ile Lys Gin Met Arg Ile Ile His Gly
50    55    60

Ser Gly Tyr Ser Asp Glu Asp Lys Arg Gly Phe Thr Lys Leu Val Tyr
65    70    75   80

Gln Asn Ile Phe Thr Ala Met Gin Ala Met Ile Asp Ala Met Asp Thr
85    90   95

Leu Lys Ile Pro Tyr Lys Tyr Glu His Arg Lys Ala His Ala Gin Leu
100   105  110

Val Arg Glu Val Asp Val Glu Lys Val Ser Ala Phe Gin Asn Pro Tyr
115   120  125

Val Asp Ala Ile Lys Ser Leu Trp Asn Asp Pro Gly Ile Gin Glu Cys
130   135  140

Tyr Asp Arg Arg Arg Glu Tyr Gin Leu Ser Asp Ser Thr Lys Tyr Tyr
145   150  155  160

Leu Asn Asp Leu Asp Arg Val Ala Asp Pro Ala Tyr Leu Pro Thr Gin
165   170  175

Gln Asp Val Leu Arg Val Arg Val Pro Thr Thr Glu Ile Ile Gin Leu
180   185  190

Pro Phe Asp Leu Gin Ser Val Ile Phe Arg Met Val Asp Val Gly Gly
195   200  205

Gln Arg Ser Glu Arg Arg Lys Trp His Cys Phe Gin Asn Val Thr
210   215  220

Ser Ile Met Phe Leu Val Ala Leu Ser Glu Tyr Asp Gin Val Leu Val
225   230  235  240

Glu Ser Asp Asn Gin Ser Met Glu Glu Ser Lys Ala Leu Phe Arg
245   250  255

Thr Ile Ile Thr Tyr Pro Trp Phe Gin Asn Ser Ser Val Ile Leu Phe
260   265  270

Leu Asn Lys Lys Asp Leu Glu Glu Lys Ile Met Tyr Ser His Leu
275   280  285

Val Asp Tyr Phe Pro Glu Tyr Asp Gly Pro Gin Arg Asp Ala Gin Ala
290   295  300

Ala Arg Glu Phe Ile Lys Met Phe Val Asp Leu Asn Pro Asp Ser
305   310  315  320

Asp Lys Ile Ile Tyr Ser His Phe Thr Cys Ala Thr Asp Thr Glu Gin
325   330  335

Ile Arg Phe Val Phe Ala Ala Val Lys Asp Thr Ile Leu Gin Leu Gin
340   345  350
Leu Lys Glu Tyr Asn Leu Val

<210> SEQ ID NO 339
<211> LENGTH: 355
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 339

Met Gly Cys Thr Val Ser Ala Glu Asp Lys Ala Ala Ala Glu Arg Ser
1    5    10   15

Lys Met Ile Asp Lys Asn Leu Arg Glu Asp Gly Glu Lys Ala Ala Arg
20  25   30

Glu Val Lys Leu Leu Leu Leu Gly Ala Gly Glu Ser Gly Lys Ser Thr
35  40   45

Ile Val Lys Gln Met Lys Ile Ile His Glu Asp Gly Tyr Ser Glu Glu
50  55   60

Glu Cys Arg Gln Tyr Arg Ala Val Val Tyr Ser Asn Thr Ile Gln Ser
65  70   75   80

Ile Met Ala Ile Val Lys Ala Met Gly Asn Leu Gln Ile Asp Phe Ala
85  90   95

Asp Pro Ser Arg Ala Asp Ala Arg Gln Leu Phe Ala Leu Ser Cys
100 105  110

Thr Ala Glu Glu Gln Gly Val Leu Pro Asp Leu Ser Gly Val Ile
115 120  125

Arg Arg Leu Trp Ala Asp His Gly Val Gln Ala Cys Phe Gly Arg Ser
130 135  140

Arg Glu Tyr Gln Leu Asp Ser Ala Ala Tyr Tyr Leu Asn Asp Leu
145 150  155  160

Glu Arg Ile Ala Gln Ser Asp Tyr Ile Pro Thr Gln Gln Asp Val Leu
165 170  175

Arg Thr Arg Val Lys Thr Thr Gly Ile Val Glu Thr His Phe Thr Phe
180 185  190

Lys Asp Leu His Phe Lys Met Phe Asp Val Gly Glu Glu Arg Ser Glu
195 200  205

Arg Lys Lys Trp Ile His Cys Phe Glu Gly Val Thr Ala Ile Ile Phe
210 215  220

Cys Val Ala Leu Ser Ala Tyr Asp Leu Val Leu Ala Glu Asp Glu Glu
225 230  235  240

Met Asn Arg Met His Glu Ser Met Lys Leu Phe Asp Ser Ile Cys Asn
245 250  255

Asn Lys Trp Phe Thr Asp Ser Ile Ile Leu Phe Leu Asn Lys Lys
260 265  270

Asp Leu Phe Glu Glu Lys Ile Thr His Ser Pro Leu Thr Ile Cys Phe
275 280  285

Pro Glu Tyr Thr Gly Ala Asn Lys Tyr Asp Glu Ala Ala Ser Tyr Ile
290 295  300

Gln Ser Lys Phe Glu Asp Leu Asn Lys Arg Lys Asp Thr Lys Glu Ile
305 310  315  320

Tyr Thr His Phe Thr Cys Ala Thr Asp Thr Lys Asn Val Gln Phe Val
325 330  335
-continued

Phe Asp Ala Val Thr Asp Val Ile Ile Lys Asn Asn Leu Lys Asp Cys
340 345 350
Gly Leu Phe
355

<210> SEQ ID NO 340
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 340

Met Ser Gly Val Val Arg Thr Leu Ser Arg Cys Leu Leu Pro Ala Glu
1 5 10 15
Ala Gly Gly Ala Arg Glu Arg Ala Gly Ser Gly Ala Arg Asp Ala
20 25 30
Glu Arg Glu Ala Arg Arg Ser Arg Asp Ile Asp Ala Leu Ala
35 40 45
Arg Glu Arg Arg Ala Val Arg Arg Leu Val Lys Ile Leu Leu Leu Gly
50 55 60
Ala Gly Glu Ser Gly Lys Ser Thr Phe Leu Lys Glu Met Arg Ile Ile
65 70 75 80
His Gly Arg Glu Phe Asp Glu Lys Ala Leu Leu Glu Phe Arg Asp Thr
85 90 95
Ile Phe Asp Asn Ile Leu Lys Gly Ser Arg Val Leu Val Asp Ala Arg
100 105 110
Asp Lys Leu Gly Ile Pro Trp Glu Tyr Ser Glu Asn Glu Lys His Gly
115 120 125
Met Phe Leu Met Ala Phe Glu Asn Ala Gly Leu Pro Val Glu Pro
130 135 140
Ala Thr Phe Glu Leu Tyr Val Pro Ala Leu Ser Ala Leu Trp Arg Asp
145 150 155 160
Ser Gly Ile Arg Glu Ala Phe Ser Arg Ser Glu Phe Glu Leu Gly
165 170 175
Glu Ser Val Lys Tyr Phe Leu Asp Asn Leu Asp Arg Ile Gly Glu Leu
180 185 190
Asn Tyr Phe Pro Ser Lys Glu Asp Ile Leu Leu Ala Arg Lys Ala Thr
195 200 205
Lys Gly Ile Val Glu His Asp Phe Val Ile Lys Ile Pro Phe Lys
210 215 220
Met Val Asp Val Gly Glu Glu Ser Glu Arg Glu Glu Gln Tyr Phe Glu
225 230 235 240
Cys Phe Asp Gly Ile Thr Ser Ile Leu Phe Met Val Ser Ser Ser Glu
245 250 255
Tyr Asp Glu Val Leu Met Glu Asp Arg Arg Thr Asn Arg Leu Val Glu
260 265 270
Ser Met Asn Ile Phe Glu Thr Ile Val Asn Asn Leu Phe Phe Asn
275 280 285
Val Ser Ile Ile Leu Phe Leu Asn Met Asp Leu Leu Val Glu Lys
290 295 300
Val Lys Thr Val Ser Ile Lys His Phe Pro Asp Phe Arg Gly Asp
305 310 315 320
Pro His Gln Leu Glu Asp Val Gln Arg Tyr Leu Val Gln Cys Phe Asp 325 330 335
Arg Lys Arg Arg Asn Arg Ser Lys Pro Leu Phe His His Phe Thr Thr 340 345 350
Ala Ile Asp Thr Glu Asn Val Arg Phe Val Phe His Ala Val Lys Asp 355 360 365
Thr Ile Leu Gln Glu Asn Leu Lys Asp Ile Met Leu Gln 370 375 380

<210> SEQ ID NO 341
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 341
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg 20 25 30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp 35 40 45
Asp Lys
50

<210> SEQ ID NO 342
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 342
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys
20 25

<210> SEQ ID NO 343
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 343
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr 1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg 20 25 30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp 35 40 45
Asp Lys
50

<210> SEQ ID NO 344
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 344

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1    5    10    15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20   25   30

Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp
35   40   45

Asp Lys
50

<210> SEQ ID NO 345
<211> LENGTH: 343
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<222> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 345

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1    5    10    15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Gly Lys
20   25   30

Leu Val Ile Trp Ile Asn Gly Asp Gly Tyr Asn Gly Leu Ala Glu
35   40   45

Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50   55   60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Glu Val Ala Ala Thr Gly
65   70   75   80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85   90

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Glu
100  105  110

Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115  120  125

Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130  135  140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145  150  155  160

Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Ann
165  170

Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180  185  190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195  200  205

Val Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210  215  220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225  230  235  240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245  250  255
-continued

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile
340

<210> SEQ ID NO 346
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 346
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20  25  30
Leu Val Ile Trp Ile Asn Gly Asp Gly Tyr Asn Gly Leu Ala Glu
35  40  45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50  55  60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65  70  75  80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85  90  95
Ala Gin Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Ann
165 170 175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
<table>
<thead>
<tr>
<th>Ala</th>
<th>Trp</th>
<th>Ser</th>
<th>Asn</th>
<th>Ile</th>
<th>Asp</th>
<th>Thr</th>
<th>Ser</th>
<th>Lys</th>
<th>Val</th>
<th>Asn</th>
<th>Tyr</th>
<th>Gly</th>
<th>Val</th>
<th>Thr</th>
<th>Val</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>260</td>
<td></td>
<td></td>
<td></td>
<td>265</td>
<td></td>
<td></td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Pro</td>
<td>Thr</td>
<td>Phe</td>
<td>Lys</td>
<td>Gly</td>
<td>Gln</td>
<td>Pro</td>
<td>Ser</td>
<td>Lys</td>
<td>Pro</td>
<td>Phe</td>
<td>Val</td>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td>280</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Ala</td>
<td>Gly</td>
<td>Ile</td>
<td>Asn</td>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td>Asn</td>
<td>Lys</td>
<td>Ala</td>
<td>Glu</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>290</td>
<td></td>
<td></td>
<td></td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>Glu</td>
<td>Asn</td>
<td>Tyr</td>
<td>Leu</td>
<td>Leu</td>
<td>Thr</td>
<td>Asp</td>
<td>Glu</td>
<td>Gly</td>
<td>Leu</td>
<td>Glu</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td>310</td>
<td></td>
<td></td>
<td></td>
<td>315</td>
<td></td>
<td></td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Lys</td>
<td>Pro</td>
<td>Leu</td>
<td>Gly</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td>Ser</td>
<td>Tyr</td>
<td>Glu</td>
<td>Glu</td>
<td>Glu</td>
</tr>
<tr>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td>330</td>
<td></td>
<td></td>
<td></td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Ile</td>
<td>Ala</td>
<td>Thr</td>
<td>Met</td>
<td>Glu</td>
<td>Asn</td>
<td>Ala</td>
<td>Gin</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>340</td>
<td></td>
<td></td>
<td>345</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Phe</td>
<td>Trp</td>
<td>Tyr</td>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Ile</td>
<td>Glu</td>
<td>Ala</td>
<td>Arg</td>
<td>Leu</td>
<td>Gin</td>
<td>Ala</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>355</td>
<td></td>
<td></td>
<td>360</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Glu</td>
<td>Ser</td>
<td>Arg</td>
<td>Asp</td>
<td>Tyr</td>
<td>Lys</td>
<td>Asp</td>
<td>Asp</td>
<td>Asp</td>
<td>Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td></td>
<td></td>
<td>375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 347
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 347

<table>
<thead>
<tr>
<th>Met</th>
<th>Lys</th>
<th>Ile</th>
<th>Lys</th>
<th>Thr</th>
<th>Gly</th>
<th>Ala</th>
<th>Arg</th>
<th>Ile</th>
<th>Leu</th>
<th>Ala</th>
<th>Leu</th>
<th>Ser</th>
<th>Ala</th>
<th>Leu</th>
<th>Thr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr</td>
<td>Met</td>
<td>Met</td>
<td>Phe</td>
<td>Ser</td>
<td>Ala</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td>Ile</td>
<td>Glu</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Val</td>
<td>Ile</td>
<td>Trp</td>
<td>Ile</td>
<td>Asn</td>
<td>Gly</td>
<td>Asp</td>
<td>Lys</td>
<td>Gly</td>
<td>Tyr</td>
<td>Asn</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Gly</td>
<td>Lys</td>
<td>Phe</td>
<td>Glu</td>
<td>Lys</td>
<td>Asp</td>
<td>Thr</td>
<td>Gly</td>
<td>Ile</td>
<td>Lys</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Pro</td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Glu</td>
<td>Lys</td>
<td>Phe</td>
<td>Pro</td>
<td>Gin</td>
<td>Val</td>
<td>Ala</td>
<td>Thr</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp</td>
<td>Gly</td>
<td>Pro</td>
<td>Asp</td>
<td>Ile</td>
<td>Phe</td>
<td>Trp</td>
<td>Ala</td>
<td>His</td>
<td>Asp</td>
<td>Arg</td>
<td>Phe</td>
<td>Gly</td>
<td>Glu</td>
<td>Tyr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala</td>
<td>Gln</td>
<td>Ser</td>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Ile</td>
<td>Thr</td>
<td>Pro</td>
<td>Asp</td>
<td>Lys</td>
<td>Ala</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp</td>
<td>Lys</td>
<td>Leu</td>
<td>Tyr</td>
<td>Pro</td>
<td>Phe</td>
<td>Thr</td>
<td>Asp</td>
<td>Ala</td>
<td>Val</td>
<td>Arg</td>
<td>Tyr</td>
<td>Asn</td>
<td>Gly</td>
<td>Lys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Tyr</td>
<td>Pro</td>
<td>Ile</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
<td>Ser</td>
<td>Leu</td>
<td>Ile</td>
<td>Tyr</td>
<td>Ann</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asp</td>
<td>Leu</td>
<td>Leu</td>
<td>Asp</td>
<td>Pro</td>
<td>Pro</td>
<td>Lys</td>
<td>Thr</td>
<td>Trp</td>
<td>Glu</td>
<td>Ile</td>
<td>Pro</td>
<td>Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Asp</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Lys</td>
<td>Ser</td>
<td>Ala</td>
<td>Leu</td>
<td>Met</td>
<td>Phe</td>
<td>Ann</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Gln</td>
<td>Glu</td>
<td>Pro</td>
<td>Tyr</td>
<td>Phe</td>
<td>Thr</td>
<td>Trp</td>
<td>Pro</td>
<td>Leu</td>
<td>Ile</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr</td>
<td>Ala</td>
<td>Phe</td>
<td>Lys</td>
<td>Tyr</td>
<td>Glu</td>
<td>Asn</td>
<td>Gly</td>
<td>Lys</td>
<td>Tyr</td>
<td>Asp</td>
<td>Ile</td>
<td>Lys</td>
<td>Gly</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Asp</td>
<td>Ann</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
<td>Lys</td>
<td>Gly</td>
<td>Leu</td>
<td>Thr</td>
<td>Phe</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Ann
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Gly Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Thr Met Glu Ann Ala Asn Ser
340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gln Ala Ser Val
355 360 365
Asp Ala Glu Ser Arg Asp Tyr Lys Asp Asp Asp Lys
370 375 380

<210> SEQ ID NO 348
<211> LENGTH: 381
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence
<400> SEQUENCE: 348

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5  10  15
Thr Met Met Phe Ser Ser Ala Ser Ala Leu Lys Ile Glu Glu Gly Lys
20 25 30
Leu Val Ile Trp Ile Asn Gly Asp Lys Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60
His Pro Asp Lys Leu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly
65 70 75 80
Asp Gly Pro Asp Ile Phe Trp Ala His Asp Arg Phe Gly Tyr
85 90 95
Ala Gin Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
105 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Ann
130 135 140
Lys Asp Leu Leu Pro Asn Pro Phe Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Ann
165 170 175
Leu Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
OTHER INFORMATION: Synthetic amino acid sequence

SEQUENCE: 350

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15
Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
20  25  30
Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro
35  40  45
Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala
50  55  60
Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile
65  70  75  80
Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala
85  90  95
Thr Lys Val Gly Ala Leu Ser Lys Gly Glu Leu Lys Glu Asn Leu Ala
100 105 110
Asp Tyr Lys Asp Asp Asp Asp Lys
115 120

SEQ ID NO 351
LENGTH: 120
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic amino acid sequence

SEQUENCE: 351

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15
Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
20  25  30
Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro
35  40  45
Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala
50  55  60
Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile
65  70  75  80
Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala
85  90  95
Thr Lys Val Gly Ala Leu Ser Lys Gly Glu Leu Lys Glu Asn Leu Ala
100 105 110
Asp Tyr Lys Asp Asp Asp Asp Lys
115 120

SEQ ID NO 352
LENGTH: 120
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic amino acid sequence

SEQUENCE: 352

Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr
1  5  10  15
Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu
-continued

Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35 40 45
Leu Thr Asp Ser Phe Asp Thr Val Leu Lys Ala Asp Gly Ala
50 95 96
Ile Leu Val Asp Phe Trp Ala Glu Trp Cyc Gly Pro Cyc Lys Met Ile
65 70 75 80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Glu Gly Lys Leu Thr
85 90 95
Val Ala Lys Leu Asn Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr
100 105 110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Phe Lys Asn Gly Glu Val
115 120 125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Glu Leu Gly Asn
130 135 140
Leu Ala Asp Tyr Lys Asp Asp Asp Gly
145 150

<210> SEQ ID NO 355
<211> LENGTH: 154
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 355

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20 25 30
Leu Gln Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35 40 45
Leu Thr Asp Ser Phe Asp Thr Val Leu Lys Ala Asp Gly Ala
50 95 96
Ile Leu Val Asp Phe Trp Ala Glu Trp Cyc Gly Pro Cyc Lys Met Ile
65 70 75 80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Glu Gly Lys Leu Thr
85 90 95
Val Ala Lys Leu Asn Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr
100 105 110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Phe Lys Asn Gly Glu Val
115 120 125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Glu Leu Gly Asn
130 135 140
Leu Ala Asp Tyr Lys Asp Asp Asp Gly
145 150

<210> SEQ ID NO 356
<211> LENGTH: 154
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 356

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Ile Glu Ala Arg
20 25 30
Leu Glu Ala Ser Val Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His
35 40 45
Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala
50 55 60
Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile
65 70 75 80
Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Glu Gly Lys Leu Thr
85 90 95
Val Ala Lys Leu Asn Ile Asp Glu Asn Pro Gly Thr Ala Pro Lys Tyr
100 105 110
Gly Ile Arg Gly Ile Pro Thr Leu Leu Phe Lys Asn Gly Glu Val
115 120 125
Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Glu Leu Lys Glu Asn
130 135 140
Leu Ala Asp Tyr Lys Asn Asp Asp Asp Lys
145 150

<210> SEQ ID NO 357
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 357
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20 25 30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60
His Pro Asp Lys Leu Glu Glu Gly Lys Phe Pro Gin Val Ala Ala Thr Gly
65 70 75 80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85 90 95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asn Pro Lys Thr Trp Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
165 170 175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Glu
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Ann
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Glu Ser
340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gln Ala Ser Val
355 360 365
Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser
370 375 380
Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
385 390 395 400
Trp Ala Glu Glu Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
405 410 415
Glu Ile Ala Asp Glu Glu Tyr Glu Gly Lys Leu Thr Val Ala Lys Leu Ann
420 425 430
Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
435 440 445
Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val
450 455 460
Gly Ala Leu Ser Lys Gly Glu Leu Lys Glu Asn Ala Asp Tyr Lys
465 470 475 480
Asp Asp Asp Asp Lys
485

<210> SEQ ID NO 358
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 358
Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1 5 10 15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Glu Lys
20 25 30
Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35 40 45
Val Gly Lys Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50 55 60
His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gin Val Ala Ala Thr Gly
65  70  75  80
Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
95  90  95
Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gin
100 105 110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140
Lys Asp Leu Leu Pro Asp Pro Pro Lys Thr Trp Glu Glu Ile Pro Ala
145 150 155 160
Leu Asp Lys Glu Leu Lys Ala Lys Gly Lys Ser Ala Leu Met Phe Asn
165 170 175
Leu Gin Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205
Val Asp Asn Ala Gly Ala Lys Leu Thr Phe Leu Val Asp Leu
210 215 220
Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240
Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255
Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270
Leu Pro Thr Phe Lys Gly Gin Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285
Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300
Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320
Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335
Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Gin Ser
340 345 350
Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gin Ala Ser Val
355 360 365
Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser
370 375 380
Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
385 390 395 400
Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
405 410 415
Glu Ile Ala Asp Glu Tyr Gin Gly Lys Leu Thr Val Ala Lys Leu Asn
420 425 430
Ile Asp Gin Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
435 440 445
Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val
450 455 460
<210> SEQ ID NO 359
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 359

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1  5    10   15
Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20   25   30
Leu Val Ile Trp Ile Asn Gly Asp Gly Tyr Asn Gly Leu Ala Glu
35   40   45
Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50   55   60
His Pro Asp Lys Leu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65   70   75   80
Asp Gly Pro Asp Ile Ile Phe Thr Ala His Asp Arg Phe Gly Gly Tyr
85   90   95
Ala Glu Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100  105  110
Asp Lys Leu Tyr Pro Phe Thr Trp Asp Ala Val Arg Tyr Asn Gly Lys
115  120  125
Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130  135  140
Lys Asp Pro Pro Leu Pro Asn Pro Pro Lys Thr Trp Glu Ile Pro Ala
145  150  155  160
Leu Asp Lys Glu Leu Lys Ala Gly Leu Ser Ala Leu Pro Phe Asn
165  170  175
Leu Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180  185  190
Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195  200  205
Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210  215  220
Ile Lys Asn His Met Asn Ala Asp Thr Tyr Ser Ile Ala Glu
225  230  235  240
 Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245  250  255
 Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260  265  270
 Leu Pro Thr Phe Lys Gly Gln Pro Ser Lys Pro Phe Val Gly Val Leu
275  280  285
 Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290  295  300
Fhe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305  310  315  320
-continued

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335

Leu Ala Lys Asp Pro Arg Ile Ala Thr Met Glu Asn Ala Gln Ser
340 345

Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Gln Ala Ser Val
355 360 365

Asp Ala Glu Ser Arg Ser Asp Lys Ile Ile His Leu Thr Asp Asp Ser
370 375 380

Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
385 390 395 400

Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
405 410 415

Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn
420 425 430

Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
435 440 445

Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Asp Thr Lys Val
450 455 460

Gly Ala Leu Ser Lys Gly Gin Leu Lys Glu Asn Leu Ala Asp Tyr Lys
465 470 475 480

Asp Asp Asp Asp Lys
485

---

<210> SEQ ID NO 360
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FRAGMENT:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 360

Met Lys Ile Lys Thr Gly Ala Arg Ile Leu Ala Leu Ser Ala Leu Thr
1   5   10  15

Thr Met Met Phe Ser Ala Ser Ala Leu Ala Lys Ile Glu Glu Gly Lys
20  25  30

Leu Val Ile Trp Ile Asn Gly Asp Lys Gly Tyr Asn Gly Leu Ala Glu
35  40  45

Val Gly Lys Phe Glu Lys Asp Thr Gly Ile Lys Val Thr Val Glu
50  55  60

His Pro Asp Lys Leu Glu Glu Lys Phe Pro Gln Val Ala Ala Thr Gly
65  70  75  80

Asp Gly Pro Asp Ile Ile Phe Trp Ala His Asp Arg Phe Gly Gly Tyr
85  90  95

Ala Gln Ser Gly Leu Leu Ala Glu Ile Thr Pro Asp Lys Ala Phe Gln
100 105 110

Asp Lys Leu Tyr Pro Phe Thr Asp Ala Val Arg Tyr Asn Gly Lys
115 120 125

Leu Ile Ala Tyr Pro Ile Ala Val Glu Ala Leu Ser Leu Ile Tyr Asn
130 135 140

Lys Asp Leu Leu Pro Asn Pro Pro Lys Thr Trp Glu Ile Pro Ala
145 150 155 160

Leu Asp Lys Glu Leu Lys Ala Gly Lys Ser Ala Leu Met Phe Ann
165 170 175
Leu Gln Glu Pro Tyr Phe Thr Trp Pro Leu Ile Ala Ala Asp Gly Gly
180 185 190

Tyr Ala Phe Lys Tyr Glu Asn Gly Lys Tyr Asp Ile Lys Asp Val Gly
195 200 205

Val Asp Asn Ala Gly Ala Lys Ala Gly Leu Thr Phe Leu Val Asp Leu
210 215 220

Ile Lys Asn Lys His Met Asn Ala Asp Thr Asp Tyr Ser Ile Ala Glu
225 230 235 240

Ala Ala Phe Asn Lys Gly Glu Thr Ala Met Thr Ile Asn Gly Pro Trp
245 250 255

Ala Trp Ser Asn Ile Asp Thr Ser Lys Val Asn Tyr Gly Val Thr Val
260 265 270

Leu Pro Thr Phe Lys Gly Glu Glu Pro Ser Lys Pro Phe Val Gly Val Leu
275 280 285

Ser Ala Gly Ile Asn Ala Ala Ser Pro Asn Lys Glu Leu Ala Lys Glu
290 295 300

Phe Leu Glu Asn Tyr Leu Leu Thr Asp Glu Gly Leu Glu Ala Val Asn
305 310 315 320

Lys Asp Lys Pro Leu Gly Ala Val Ala Leu Lys Ser Tyr Glu Glu Glu
325 330 335 340

Leu Ala Lys Asp Pro Arg Ile Ala Ala Thr Met Glu Asn Ala Glu Ser
345 350

Ala Phe Trp Tyr Ala Val Arg Ile Glu Ala Arg Leu Glu Ala Ser Val
360 365

Asp Ala Glu Ser Arg Ser Asp Lys Ile His Leu Thr Asp Asp Ser
370 375 380

Phe Asp Thr Asp Val Leu Lys Ala Asp Gly Ala Ile Leu Val Asp Phe
385 390 395 400

Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
405 410 415

Glu Ile Ala Asp Glu Tyr Glu Gly Lys Leu Thr Val Ala Lys Leu Asn
420 425 430

Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile
435 440 445

Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala Ala Thr Lys Val
450 455 460

Gly Ala Leu Ser Lys Gly Glu Leu Gly Asn Leu Ala Asp Tyr Lys
465 470 475 480

Asp Asp Asp Asp Asp Lys
485

<210> SEQ ID NO 361
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Synthetic nucleic acid sequence
<400> SEQUENCE: 361
atgcctaaag cogtccctc a
21
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 362

atgagccgc tcct 15
<210> SEQ ID NO 363
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 363

augagccgc ucucc 15
<210> SEQ ID NO 364
<211> LENGTH: 5
<212> TYPE: PRO
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 364

Met Pro Lys Ala Ala
1  5

<210> SEQ ID NO 365
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 365

tggttacctt atctc 15
<210> SEQ ID NO 366
<211> LENGTH: 5
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic nucleic acid sequence

<400> SEQUENCE: 366

uggguaucuc acuca 15

<210> SEQ ID NO 367
<211> LENGTH: 5
<212> TYPE: PRO
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic amino acid sequence

<400> SEQUENCE: 367

Trp Val Thr His Ser
1  5

<210> SEQ ID NO 368
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
What is claimed is:

1. A fully intact eubacterial minicell derived from a eubacterial parent cell, wherein the minicell comprises a biologically active compound and displays Invasin, wherein the biologically active compound is exogenous to the parent cell and distinct from Invasin.

2. The minicell of claim 1, wherein the Invasin comprises a Yersinia invasin or a fragment thereof.

3. The minicell of claim 1, wherein the Invasin comprises the invasin protein from Yersinia pseudotuberculosis or a fragment thereof.

4. The minicell of claim 1, wherein the biologically active compound is selected from the group consisting of a polypeptide, a small molecule, and a nucleic acid.

5. The minicell of claim 4, wherein the biologically active compound is a polypeptide.

6. The minicell of claim 5, wherein the polypeptide is a therapeutic polypeptide.

7. The minicell of claim 6, wherein the therapeutic polypeptide is a soluble polypeptide.

8. The minicell of claim 6, wherein the therapeutic polypeptide is a fusion protein.

9. The minicell of claim 6, wherein the therapeutic polypeptide is a protein toxin.

10. The minicell of claim 4, wherein the biologically active compound is a small molecule.

11. The minicell of claim 10, wherein the small molecule is a drug molecule.

12. The minicell of claim 4, wherein the biologically active compound is a nucleic acid.

13. The minicell of claim 12, wherein the nucleic acid is a therapeutic nucleic acid.

14. The minicell of claim 13, wherein the therapeutic nucleic acid comprises an antisense nucleic acid.

15. The minicell of claim 13, wherein the therapeutic nucleic acid comprises a ribozyme.

16. The minicell of claim 13, wherein the therapeutic nucleic acid is comprises an aptamer.

* * * * *