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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 April 2009 (02.04.2009)

PCT

(10) International Publication Number
WO 2009/041831 A1

(51) International Patent Classification:
C07K 14/01 (2006.01) *C12N 7/01* (2006.01)
C07H 21/04 (2006.01)

(21) International Application Number:
PCT/NZ2008/000248

(22) International Filing Date:
25 September 2008 (25.09.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/975,104	25 September 2007 (25.09.2007)	US
60/989,840	22 November 2007 (22.11.2007)	US
60/989,841	22 November 2007 (22.11.2007)	US

(71) Applicant *(for all designated States except US)*: **PASTORAL GREENHOUSE GAS RESEARCH LIMITED** [NZ/NZ]; Level 13, 113-119 The Terrace, Wellington, 6011 (NZ).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ALTERMANN, Eric Heinz** [DE/NZ]; 26 Akaroa Avenue, Palmerston North (NZ). **ATTWOOD, Graeme Trevor** [NZ/NZ]; 136 Pohangina Valley East Road, Rd14, Ashhurst (NZ). **LEAHY, Sinead Christine** [IE/NZ]; 1 Moana Street, Hokuwhitu, Palmerston North (NZ). **KELLY, William John** [NZ/NZ]; 208 Cambridge Avenue, Ashhurst (NZ). **RONIMUS, Robert Starr** [NZ/NZ]; 468 Stoney Creek Road, Rd10, Palmerston North (NZ). **DONG, Li** [NZ/NZ]; 29 Kingston Street, Awapuni, Palmerston North (NZ). **KONG, Zhanhao** [NZ/CN]; 45-3-501 Liuying Road, Shanghai (CN). **SCHOFIELD, Linley Rose** [NZ/NZ]; 3 Slacks Road, Palmerston North (NZ). **DEY, Debjit** [NZ/NZ]; 6 Waterford Place, Palmerston North (NZ). **TOOTILL, Catherine Mary** [NZ/NZ]; 22 Balmoral Drive, Palmerston North (NZ). **SANG, Carrie** [CN/NZ]; 29 Swansea Street, Palmerston North (NZ). **MOON, Christina Diane** [NZ/NZ]; 10 Chester Crescent, Palmerston North (NZ). **PETRUS, Hendricus, Janssen** [NZ/NZ]; 155 A Victoria Avenue, Palmerston North (NZ).

(74) Agent: **BALDWINS**; P O Box 852, Wellington (NZ).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *with international search report*

(54) **Title:** PHAGE φ-MRU POLYNUCLEOTIDES AND POLYPEPTIDES AND USES THEREOF

(57) **Abstract:** The invention encompasses phage φmru including phage induction, phage particles, and the phage genome. Also encompassed are phage polypeptides, as well as polynucleotides which encode these polypeptides, expression vectors comprising these polynucleotides, and host cells comprising these vectors. The invention further encompasses compositions and methods for detecting, targeting, permeabilising, and inhibiting microbial cells, especially methanogen cells, using the disclosed phage, polypeptides, polynucleotides, expression vectors, or host cells.

WO 2009/041831 A1

PHAGE φmru POLYNUCLEOTIDES AND POLYPEPTIDES AND USES THEREOF

RELATED APPLICATIONS

This application claims the benefit of U.S. Application No. 60/975,104, filed September 25, 2007, U.S. Application No. 60/989,840, filed November 22, 2007, and U.S. Application No. 60/989,841, filed November 22, 2007, the contents of all of which are hereby incorporated herein by reference in their entirety.

FIELD OF THE INVENTION

The invention relates to compositions and methods for delivering inhibitory molecules into microbial cells, in particular, methanogen cells. Specifically, the invention relates to newly identified phage φmru, including phage induction, phage particles, and the phage genome, and also phage polypeptides, as well as polynucleotides which encode these polypeptides. The invention also relates to expression vectors and host cells for producing these polypeptides. The invention further relates to methods for detecting, targeting, and inhibiting microbial cells, especially methanogen cells, using the disclosed phage, polypeptides, polynucleotides, expression vectors, and host cells.

BACKGROUND OF THE INVENTION

In New Zealand, agricultural activity accounts for the majority of greenhouse gas emissions. Therefore, reducing agricultural emissions of greenhouse gases is important for meeting New Zealand's obligations under the Kyoto Protocol. The Protocol requires reduction of greenhouse gases to 1990 levels by the end of the first commitment period (2008-2012). To this end, agricultural sector groups and the New Zealand government established the Pastoral Greenhouse Gas Research Consortium (PGGRC) to identify means for reducing New Zealand's agricultural greenhouse gas emissions.

An important part of the PGGRC's activities has been research into reducing methane emissions from New Zealand's grazing ruminants. Mitigating methane emissions from ruminants is of commercial interest for two reasons. First, failure to meet commitments under the Kyoto Protocol will force the government to purchase carbon credits. This is currently estimated to cost $350 million. Second, methane production results in the loss of 8-12% of the gross energy produced in the rumen. This energy could be used, instead, to improve ruminant productivity.

Methane is produced in the rumen by microbes called methanogens which are part of the phylum *Euryarchaeota* within the kingdom *Archaea*. Most methanogens grow on

CO_2 and H_2 as their sole energy source, but some can use acetate or methyl compounds for growth. Several different genera of methanogenic archaea exist in the rumen, but species of the genus *Methanobrevibacter*, especially *M. ruminantium*, and *M. smithii* are thought to be the predominant methanogens in New Zealand ruminants. *M. ruminantium* is currently the subject of a genome sequencing project funded by the PGGRC. The project is the first genome sequencing of a rumen methanogen and it aims to build a better understanding of the biology of *Methanobrevibacter* to discover targets for inhibition of methane formation.

Reducing methane production in the rumen requires the inhibition of methanogens or the inactivation of their methanogenesis pathway. A means of inhibiting methane production is to deliver specific inhibitory molecules into methanogen cells. This may be achieved, for example, by use of agents, such as bacteriophage, which specifically target methanogens. Several phage have been characterised for non-rumen methanogens but there have been no published accounts of phage able to infect or lyse rumen methanogens. Therefore, it would be highly advantageous to identify phage that have the ability to infect methanogen cells and/or deliver inhibitors.

SUMMARY OF THE INVENTION

The invention features an isolated phage φmru, including a phage particle and/or phage genome, produced in whole or in part, as well as isolated polynucleotides and polypeptides of the phage as described in detail herein.

The invention also features an isolated polypeptide comprising at least one phage amino acid sequence selected from the group consisting of SEQ ID NO:1-69. In a particular aspect, the polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NO:2-5 and 62-68. In a further aspect, the polypeptide comprises the amino acid sequence of SEQ ID NO:63. In another aspect, the polypeptide is a fragment, for example, comprising at least one amino acid sequence extending from residues 32-186 of SEQ ID NO:63.

The invention additionally features an isolated polynucleotide comprising a coding sequence for at least one phage polypeptide. In one aspect, the polynucleotide comprises a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69. In a particular aspect, the polynucleotide comprises a coding sequence for a sequence selected from the group consisting of SEQ ID NO:2-5 and 62-68. In a further aspect, the polynucleotide comprises a coding

sequence for SEQ ID NO:63. In another aspect, the polynucleotide comprises a fragment of a coding sequence, for example, least one amino acid sequence extending from residues 32-186 of SEQ ID NO:63.

In an additional aspect, the invention features an isolated polynucleotide comprising a phage nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142. In a particular aspect, the polynucleotide comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO:75-78 and 135-141, or is particularly, SEQ ID NO:136. In another aspect, the polynucleotide is a fragment or an oligonucleotide comprising, for example, the nucleic acid sequence extending from nucleotides 94-558 of SEQ ID NO:136. In addition, the invention encompasses an isolated polynucleotide, or fragment thereof, which hybridizes to any one of the nucleic acid sequences of SEQ ID NO:74-142. The invention further encompasses an isolated polynucleotide comprising the complement, reverse complement, reverse sequence, or fragments thereof, of any one of the nucleic acid sequences.

The invention features an expression vector comprising a polynucleotide comprising a coding sequence for at least one phage polypeptide. In one aspect, the expression vector comprises a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69. In a particular aspect, the expression vector comprises a coding sequence for at least one amino acid sequence of SEQ ID NO:2-5 and 62-68. In a further aspect, the expression vector comprises a coding sequence for at least one amino acid sequence of SEQ ID NO:63. In another aspect, the expression vector comprises a coding sequence for at least one amino acid sequence extending from residues 32-186 of SEQ ID NO:63.

As a specific aspect, the invention features an expression vector which produces phage φmru, in whole or in part, as described in detail herein. In particular, the expression vector may produce phage particles, a phage genome, or modified phage, including any alterations, derivatives, variants, or fragments thereof.

The invention also features a host cell, for example, a microbial host cell, comprising at least one expression vector.

The invention specifically features an antibody raised to a polypeptide or polynucleotide as disclosed herein. In certain aspects, the antibody is directed to at least one polypeptide sequence selected from the group consisting of SEQ ID NO:1-69, or a modified sequence thereof. In alternate aspects, the antibody is raised to at least a fragment of a polynucleotide selected from the group consisting of SEQ ID NO:74-142, or a complement, or modified sequence thereof. In another aspect, the antibody includes one or more fusions or conjugates with at least one cell inhibitor, for example, anti-methanogenesis compounds (e.g., bromoethanesulphonic acid), antibodies and antibody fragments, lytic enzymes, peptide nucleic acids, antimicrobial peptides, and other antibiotics as described in detail herein.

The invention additionally features modified phage polypeptides, e.g., for at least one of SEQ ID NO:1-69, including biologically active alterations, fragments, variants, and derivatives, described herein. The invention additionally features modified antibodies, e.g., directed to at least one of SEQ ID NO:1-69, including biologically active alterations, fragments, variants, and derivatives, described herein. Also featured are polynucleotides encoding these modified polypeptides, as well as alterations, fragments, variants, and derivatives of the disclosed polynucleotides, expression vectors comprising these polynucleotides, and host cells comprising these vectors. In specific aspects, the compositions and methods of the invention employ these modified polynucleotides or polypeptides, or corresponding expression vectors or host cells.

In addition, the invention features phage polypeptides, e.g., at least one of SEQ ID NO:1-69 or modified sequences thereof, which include fusions or conjugates with at least one cell inhibitor, for example, anti-methanogenesis compounds (e.g., bromoethanesulphonic acid), antibodies and antibody fragments, lytic enzymes, peptide nucleic acids, antimicrobial peptides, and other antibiotics as described in detail herein.

The invention features a composition comprising an isolated polypeptide, e.g., at least one of SEQ ID NO:1-69, or a modified sequence thereof. The invention additionally features a composition comprising an antibody, e.g., directed to at least one of SEQ ID NO:1-69, or a modified sequence thereof. Also featured is a composition comprising an isolated polynucleotide, e.g., at least one of SEQ ID NO:74-142, or a complement or modified sequence thereof. Further featured is a composition that includes an expression vector, or host cell comprising an expression vector, in accordance with the invention. The composition can include any one of the biologically active alterations,

fragments, variants, and derivatives described herein. The compositions can include at least one cell inhibitor (e.g., as a fusion or conjugate), and can be formulated, for example, as pharmaceutical compositions or as food supplements, in particular, ruminant feed components.

The invention also features a composition of the invention as part of a kit for targeting and/or inhibiting microbial cells, especially methanogen cells, in accordance with the disclosed methods. The kits comprise: a) at least one composition as set out herein; and b) optionally, instructions for use, for example, in targeting cells or inhibiting cell growth or replication for methanogens or other microbes.

The invention features a method for producing a phage, the method comprising: a) culturing an expression vector or host cell comprising an expression vector, which comprises at least part of the phage genome under conditions suitable for the production of the phage; and b) recovering the phage from the culture. In particular aspects, the phage comprises at least one polypeptide selected from the group consisting of SEQ ID NO:1-69, or modified sequences thereof. In further aspects, the phage comprises at least one polynucleotide selected from the group consisting of SEQ ID NO:74-142, or modified sequences thereof.

The invention also features a method for producing a phage polypeptide, the method comprising: a) culturing an expression vector or host cell comprising an expression vector, which comprises at least part of a coding sequence for at least one phage polypeptide under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the culture. In particular aspects, the polypeptide comprises at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or modified sequences thereof.

The invention additionally features a method for producing a phage polypeptide, e.g., for at least one of SEQ ID NO:1-69, which comprises a fusion or conjugate with at least one cell inhibitor, for example, anti-methanogenesis compounds (e.g., bromoethanesulphonic acid), antibodies and antibody fragments, lytic enzymes, peptide nucleic acids, antimicrobial peptides, and other antibiotics as described in detail herein. Such method comprises: a) culturing an expression vector or host cell comprising an expression vector, which comprises a coding sequence for at least one phage polypeptide under conditions suitable for the expression of the polypeptide; b) forming

the phage fusion or conjugate (e.g., by expression of the fused sequence or chemical conjugation to the cell inhibitor); and c) recovering the fusion or conjugate. In particular aspects, the polypeptide comprises at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or modified sequences thereof.

In addition, the invention features a method of inhibiting (e.g., inhibiting growth or replication) of a microbial cell, in particular, a methanogen cell, comprising: a) optionally, producing or isolating at least one phage polypeptide; and b) contacting the cell with the phage polypeptide. In a particular aspect, the polypeptide comprises at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or a modified sequence thereof.

As an added feature, the invention encompasses a method of inhibiting (e.g., inhibiting growth or replication) of a microbial cell, in particular, a methanogen cell, comprising: a) optionally, producing or isolating at least one phage polypeptide, which further comprises at least one cell inhibitor; and b) contacting the cell with the phage polypeptide. In a particular aspect, the polypeptide comprises at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or a modified sequence thereof.

The invention also features a method of detecting and/or measuring the levels of a phage, or a corresponding phage polypeptide or polynucleotide, comprising: 1) contacting a sample from a subject with an antibody raised to a phage polypeptide (e.g., at least one of SEQ ID NO:1-69, or a modified sequence thereof) or a corresponding polynucleotide; and 2) determining the presence or levels of the antibody complex formed with the polypeptide or polynucleotide in the sample. Such methods can also be used for detecting and/or measuring the levels of a microbial cell, in particular, a methanogen cell.

The invention features, as well, a method of detecting and/or measuring the levels of a phage, or a corresponding phage polynucleotide (e.g., a phage coding sequence), comprising: 1) contacting a sample from a subject with a complementary polynucleotide (e.g., a sequence complementary to any one of SEQ ID NO:74-142, or modified sequence thereof); and 2) determining the presence or levels of the hybridization complex formed with the phage polynucleotide in the sample. Such methods can also be used for detecting and/or measuring the levels of a microbial cell, in particular, a methanogen cell.

In particular aspects, the methods of the invention utilize *in vivo* or *in vitro* expression components. In other aspects, the methods employ polypeptides produced by recombinant, synthetic, or semi-synthetic means, or polypeptides produced by endogenous means.

Other aspects and embodiments of the invention are described herein below.

BRIEF DESCRIPTION OF THE DRAWINGS

This invention is described with reference to specific embodiments thereof and with reference to the figures.

FIGS. 1A-1B. *M. ruminantium* prophage φmru showing putative integration site sequences *att*L and *att*R (FIG. 1A), and predicted phage functional modules and gene structure (FIG. 1B).

Fig 1C: Phage induction using sterile air (oxygen stress).

Fig 1D: Initial phage induction using MitomycinC.

Fig 1E: Agarose gel electrophoresis of PCR amplicons of induced (oxygen challenge) and uninduced *M. ruminantium*. Lanes 2 and 4: 1 kb DNA marker ladder by Invitrogen. Lanes 1 and 3 represent PCRs using primer-pair R1F - L2R on DNA isolated from induced and uninduced *M. ruminantium* cultures, respectively.

FIG. 2. Prophage φmru open reading frame annotation and comments.

FIG. 3. Prophage φmru open reading frame annotation, predicted function, and comments.

FIGS. 4A-4B. *M. ruminantium* prophage φmru sequence information, including coding sequences of phage φmru (FIG. 4A), and amino acid sequences of phage φmru (FIG. 4B).

FIG. 5. Sequence alignment of phage φmru ORF 2058 with PeiP from *M. marburgensis* and PeiW from *M. wolfeii*.

FIG. 6: Protein sequence logo of signal peptide sequences from *M. ruminantium* created using LogoBar, showing the core consensus signal.

FIG. 7: Inhibitory effect of ORF 2058 on resting *M. ruminantium* cells.

FIG. 8: Inhibitory effect of ORF 2058 on *M. ruminantium* cell growth and methane production.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

"Altered" nucleic acid sequences encoding phage polypeptides, as used herein, include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that preferably encodes the same or functionally equivalent polypeptides. The encoded polypeptide or antibody may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity (e.g., cell association, cell permeabilisation, or cell lysis) or immunological activity (e.g., one or more antibody binding sites) of the polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence", as used herein, refers to an oligopeptide, peptide, polypeptide, or protein sequence, and any fragment thereof, and to naturally occurring, recombinant, synthetic, or semi-synthetic molecules. The sequences of the invention comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 150, 200, 250 amino acids, preferably at least 5 to 10, 10 to 20, 20 to 30, 30 to 40, 40 to 50, 50 to 100, 100 to 150, 150 to 200, or 200 to 250, or 250 to 4000 amino acids, and, preferably, retain the biological activity (e.g., cell association, cell permeabilisation, or cell lysis) or the immunological activity (e.g., one or more antibody binding sites) of the original sequence. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring polypeptide molecule, amino acid sequence, and like terms, are not meant to limit the

amino acid sequence to the complete, original amino acid sequence associated with the full-length molecule.

"Amplification", as used herein, refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C. W. and G. S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "antibody" should be understood in the broadest possible sense and is intended to include intact monoclonal antibodies and polyclonal antibodies. It is also intended to cover fragments and derivatives of antibodies so long as they exhibit the desired biological activity. Antibodies encompass immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. These include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fc, Fab, Fab', and Fab_2 fragments, and a Fab expression library.

Antibody molecules relate to any of the classes IgG, IgM, IgA, IgE, and IgD, which differ from one another by the nature of heavy chain present in the molecule. These include subclasses as well, such as IgG1, IgG2, and others. The light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all classes, subclasses, and types. Also included are chimeric antibodies, for example, monoclonal antibodies or fragments thereof that are specific to more than one source, e.g., one or more mouse, human, or ruminant sequences. Further included are camelid antibodies or nanobodies. It will be understood that each reference to "antibodies" or any like term, herein includes intact antibodies, as well as any fragments, alterations, derivatives, or variants thereof.

The terms "biologically active" or "functional," as used herein, refer to a polypeptide retaining one or more structural, immunological, or biochemical functions (e.g., cell association, cell permeabilisation, or cell lysis) sequence.

The terms "cell inhibitor" or "inhibitor," as used herein, refer to agents that decrease or block the growth or replication of microbial cells, especially methanogen cells. A cell inhibitor can act to decrease or block, for example, cellular division. An inhibitor can decrease or block, for example, DNA synthesis, RNA synthesis, protein synthesis, or

post-translational modifications. An inhibitor can also decrease or block the activity of enzymes involved in the methanogenesis pathway. An inhibitor can also target a cell for recognition by immune system components. Inhibition of a cell also includes cell killing and cell death, for example, from lysis, apoptosis, necrosis, etc. Useful inhibitors include, but are not limited to, anti-methanogenesis compounds (e.g., bromoethanesulphonic acid), antibodies and antibody fragments, lytic enzymes, peptide nucleic acids, antimicrobial peptides, and other antibiotics as described in detail herein.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For the sequence A-G-T, the complementary sequence is T-C-A, the reverse complement is A-C-T and the reverse sequence is T-G-A. Complementarity between two single-stranded molecules may be partial, in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding a phage polypeptide, or a nucleic acid complementary thereto. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. In preferred aspects, a nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains one or more biological function (e.g., cell association, cell permeabilisation, or cell lysis) or immunological function of the sequence from which it was derived.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology (i.e., I identity) or complete homology (i.e., 100% identity). A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization

probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

A "methanogen," as used herein, refers to microbes that produce methane gas, which include *Methanobrevibacter*, *Methanothermobacter*, *Methanomicrobium*, *Methanobacterium*, and *Methanosarcina*. Specific methanogens include, but are not limited to, *Methanobrevibacter ruminantium*, *Methanobrevibacter smithii*, *Methanobrevibacter acididurans*, *Methanobrevibacter thaueri*, *Methanobacterium bryantii*, *Methanobacterium formicicum*, *Methanothermobacter marburgensis*, *Methanothermobacter wolfeii*, *Methanosphaera stadtmanae*, *Methanomicrobium mobile*, *Methanosarcina barkeri*, *Methanosarcina mazei*, *Methanococcoides burtonii*, and *Methanolobus taylorii*. All methanogen genera and species are encompassed by this term.

"Microbial" cells as used herein, refers to naturally-occurring or genetically modified microbial cells including archaebacteria such as methanogens, halophiles, and thermoacidophiles, and eubacteria, such as cyanobacteria, spirochetes, proteobacteria, as well as gram positive and gram negative bacteria.

The term "modified" refers to altered sequences and to sequence fragments, variants, and derivatives, as described herein.

"Nucleic acid sequence" or "nucleotide sequence" as used herein, refers to a sequence of a polynucleotide, oligonucleotide, or fragments thereof, and to DNA or RNA of natural, recombinant, synthetic, or semi-synthetic origin which may be single or double stranded, and can represent the sense or antisense strand, and coding or non-coding regions. The sequences of the invention most preferably include polypeptide coding

sequences that comprise at least 12, 15, 30, 45, 60, 75, 90, 105, 120, 135, 150, 300, 450, 600, 750 nucleotides, preferably at least 15 to 30, 30 to 60, 60 to 90, 90 to 120, 120 to 150, 150 to 300, 300 to 450, 450 to 600, or 600 to 750 nucleotides, or at least 1000 nucleotides, or at least 1500 nucleotides. It will be understood that each reference to a "nucleic acid sequence" or "nucleotide sequence" herein, will include the original, full-length sequence, as well as any complements, fragments, alterations, derivatives, or variants, thereof.

The term "oligonucleotide" refers to a nucleic acid sequence comprising at least 6, 8, 10, 12, 15, 18, 21, 25, 27, 30, or 36 nucleotides, or at least 12 to 36 nucleotides, or at least 15 to 30 nucleotides, which can be used, for example, in PCR amplification, sequencing, or hybridization assays. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers," "primers," "oligomers," "oligos," and "probes," as commonly defined in the art.

"Polypeptide," as used herein, refers to the isolated polypeptides of the invention obtained from any species, preferably microbial, from any source whether natural, synthetic, semi-synthetic, or recombinant. Specifically, a phage polypeptide can be obtained from methanogen cells, such as *Methanobrevibacter* cells, in particular, *M. ruminantium*, or *M. smithii* cells. For recombinant production, a polypeptide of the invention can be obtained from microbial or eukaryotic cells, for example, *Escherichia*, *Streptomyces*, *Bacillus*, *Salmonella*, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, or plant cells. It will be understood that each reference to a "polypeptide," herein, will include the original, full-length sequence, as well as any fragments, alternations, derivatives, or variants, thereof.

The term "polynucleotide," when used in the singular or plural, generally refers to any nucleic acid sequence, e.g., any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. This includes, without limitation, single and double stranded DNA, DNA including single and double- stranded regions, single and double stranded RNA, and RNA including single and double stranded regions, hybrid molecules comprising DNA and RNA that may be single stranded or, more typically, double stranded or include single and double stranded regions. Also included are triple-stranded regions comprising RNA or DNA or both RNA and DNA. Specifically included are mRNAs, cDNAs, and genomic DNAs, and any fragments thereof. The term includes DNAs and RNAs that contain one or more

modified bases, such as tritiated bases, or unusual bases, such as inosine. The polynucleotides of the invention can encompass coding or non-coding sequences, or sense or antisense sequences, or iRNAs such as siRNAs. It will be understood that each reference to a "polynucleotide" or like term, herein, will include the full length sequences as well as any complements, fragments, alterations, derivatives, or variants thereof.

"Peptide nucleic acid" or "PNA" as used herein, refers to an antisense molecule or anti-gene agent which comprises bases linked via a peptide backbone.

The term "ruminant," as used herein, refers to animals that have a rumen as a special type of digestive organ. Ruminants include, but are not limited to, cattle, sheep, goats, buffalo, moose, antelope, caribou, and deer.

The terms "stringent conditions" or "stringency," as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "subject" includes human and non-human animals. Non-human animals include, but are not limited to, birds and mammals, such as ruminants, and in particular, mice, rabbits, cats, dogs, pigs, sheep, goats, cows, and horses.

The terms "substantially purified" or "isolated" as used herein, refer to nucleic or amino acid sequences that are removed from their cellular, recombinant, or synthetic

environment, and are at least 60% free, preferably 75% free, and most preferably at least 90% free or at least 99% free from other components with which they are associated in a cellular, recombinant, or synthetic environment.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of a polypeptide, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. A variant polynucleotide is altered by one or more nucleotides. A variant may result in "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may result in "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, LASERGENE software (DNASTAR).

The invention also encompasses variants which retain at least one biological activity (e.g., cell association, cell permeabilisation, or cell lysis) or immunological activity of the polypeptide. A preferred variant is one having substantially the same or a functionally equivalent sequence, for example, at least 80%, and more preferably at least 90%, sequence identity to a disclosed sequence. A most preferred variant is one having at least 95%, at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or at least 99.9% sequence identity to a sequence disclosed herein. The percentage identity is determined by aligning the two sequences to be compared as described below, determining the number of identical residues in the aligned portion, dividing that number

by the total number of residues in the inventive (queried) sequence, and multiplying the result by 100. A useful alignment program is AlignX (Vector NTI).

Description of the invention

Methane is produced in the foregut of ruminants by methanogens which act as terminal reducers of carbon in the rumen system. The multi-step methanogenesis pathway is well elucidated, mainly from the study of non-rumen methanogens, but the adaptations that allow methanogens to grow and persist in the rumen are not well understood. *Methanobrevibacter ruminantium* is a prominent methanogen in New Zealand ruminants. As described herein, the genome of *M. ruminantium* has been sequenced and shown as approximately 3.0 Mb in size with a GC content of 33.68%. As an unexpected finding, the *M. ruminantium* genome was found to include a prophage sequence (designated φmru) with distinct functional modules encoding phage integration, DNA replication and packaging, capsid proteins, lysis, and lysogenic conversion functions.

The *M. ruminantium* phage was identified during high through-put sequencing, when a 30 to 40 kb region of the genome was found to be over-represented in the sequenced clones. This suggested that a part of the genome was present in higher copy number than normal, and could be attributed to the replication of a resident phage. The over-represented region was investigated and detailed bioinformatic analyses of the predicted open reading frames present indicated that it contained phage-like genes. A low GC region found at the distal end of the phage sequence (lysogenic conversion) has been shown to harbour a predicted DNA modification system by sulphur (dnd) which might provide additional modification of host or foreign DNA. The *M. ruminantium* prophage sequence is described in detail herein. In various aspects of the invention, the prophage polynucleotides and polypeptides can be used as a means for inhibiting methanogens and/or methanogenesis in the rumen, and to further elucidate the role of *M. ruminantium* in methane formation.

The invention therefore encompasses phage polypeptides, including those comprising at least one of SEQ ID NO:1-69, and fragments, variants, and derivatives thereof. The invention also encompasses the use of these polypeptides for targeting and inhibiting microbial cells, especially methanogen cells. The invention further encompasses the use of the polypeptides for the inhibition of growth or replication of such cells. The polypeptides of the present invention may be expressed and used in various assays to determine their biological activity. The polypeptides may be used for large-scale

synthesis and isolation protocols, for example, for commercial production. Such polypeptides may be used to raise antibodies, to isolate corresponding amino acid sequences, and to quantitatively determine levels of the amino acid sequences. The polypeptides of the present invention may also be used as compositions, for example, pharmaceutical compositions, and as food supplements, e.g., ruminant feed components. The polypeptides of the present invention also have health benefits. In heath-related aspects, inhibitors of methanogens can be used to restore energy to the subject that is normally lost as methane. In particular aspects, slow-release ruminal devices can be used in conjunction with the polypeptides, and compositions (e.g., pharmaceutical compositions and food supplements) of the invention.

The polypeptides of the present invention comprise at least one sequence selected from the group consisting of: (a) polypeptides comprising at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or fragments, variants, or derivatives thereof; (b) polypeptides comprising a functional domain of at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, and fragments and variants thereof; and (c) polypeptides comprising at least a specified number of contiguous residues of at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or variants or derivatives thereof. In one embodiment, the invention encompasses an isolated polypeptide comprising the amino acid sequence of at least one of SEQ ID NO:1-69. All of these sequences are collectively referred to herein as polypeptides of the invention.

The invention also encompasses polynucleotides that encode at least one phage polypeptide, including those of SEQ ID NO:1-69, and fragments, variants, and derivatives thereof. The invention also encompasses the use of these polynucleotides for preparing expression vectors and host cells for targeting and inhibiting microbial cells, especially methanogen cells. The invention further encompasses the use of the polynucleotides for the inhibition of growth or replication of such cells. The isolated polynucleotides of the present invention also have utility in genome mapping, in physical mapping, and in cloning of genes of more or less related phage. Probes designed using the polynucleotides of the present invention may be used to detect the presence and examine the expression patterns of genes in any organism having sufficiently homologous DNA and RNA sequences in their cells, using techniques that are well known in the art, such as slot blot techniques or microarray analysis. Primers designed using the polynucleotides of the present invention may be used for

sequencing and PCR amplifications. The polynucleotides of the present invention may also be used as compositions, for example, pharmaceutical compositions, and as food supplements, e.g., ruminant feed components. The polynucleotides of the present invention also have health benefits. For such benefits, the polynucleotides can be presented as expression vectors or host cells comprising expression vectors. In particular aspects, slow-release ruminal devices can be used in conjunction with the polynucleotides, vectors, host cells, and compositions (e.g., pharmaceutical compositions and food supplements) of the invention.

The polynucleotides of the present invention comprise at least one sequence selected from the group consisting of: (a) sequences comprising a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or fragments or variants thereof; (b) complements, reverse sequences, and reverse complements of a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or fragments or variants thereof; (c) open reading frames contained in the coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, and their fragments and variants; (d) functional domains of a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, and fragments and variants thereof; and (e) sequences comprising at least a specified number of contiguous residues of a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69, or variants thereof. In one embodiment, the invention encompasses an isolated polynucleotide comprising a coding sequence for at least one amino acid sequence selected from the group consisting of SEQ ID NO:1-69.

The polynucleotides of the present invention comprise at least one sequence selected from the group consisting of: (a) sequences comprising at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142, or fragments or variants thereof; (b) complements, reverse sequences, and reverse complements of a coding sequence for at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142, or fragments or variants thereof; (c) open reading frames contained in the nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142, and their fragments and variants; (d) functional domains of a coding sequence of at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142, and fragments and variants thereof; and (e)

sequences comprising at least a specified number of contiguous residues of at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:74-142, or variants thereof. Oligonucleotide probes and primers and their variants obtained from any of the disclosed sequences are also provided. All of these polynucleotides and oligonucleotides are collectively referred to herein, as polynucleotides of the invention.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding the polypeptides of the invention, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard bacterial triplet genetic code as applied to naturally occurring amino acid sequences, and all such variations are to be considered as being specifically disclosed.

Nucleotide sequences which encode the phage polypeptides, or their fragments or variants, are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring sequence under appropriately selected conditions of stringency. However, it may be advantageous to produce nucleotide sequences encoding a polypeptide, or its fragment or derivative, possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the polypeptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. For example, codons can be optimized for expression in *E. coli* in accordance with known methods. Other reasons for substantially altering the nucleotide sequence encoding polypeptides and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof, which encode the polypeptides, or their fragments or variants, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding a polypeptide, or any variants or fragment thereof. Also

encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:74-149, or their complements, under various conditions of stringency as taught in Wahl, G. M. and S. L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (U.S. Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase Amersham Pharmacia Biotech (Piscataway, NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Life Technologies (Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer), or the Genome Sequencer 20™ (Roche Diagnostics).

The nucleic acid sequences encoding the polypeptides may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements, and downstream elements such as terminators and non-coding RNA structures. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Another useful method is inverse PCR, also called IPCR (see, e.g., Ochman H, Gerber AS, Hartl DL. Genetics. 1988 Nov;120(3):621-3). Inverse PCR can be employed when only one internal sequence of the target DNA is known. The inverse PCR method includes a series of digestions and self-ligations with the DNA being cut by a restriction endonuclease. This cut results in a known sequence at either end of unknown sequences. In accordance with this method, target DNA is lightly cut into smaller

fragments of several kilobases by restriction endonuclease digestion. Self-ligation is then induced under low concentrations causing the phosphate backbone to reform and produce a circular DNA ligation product. Target DNA is then restriction digested with a known endonuclease. This generates a cut within the known internal sequence generating a linear product with known terminal sequences. This product can then be used for standard PCR conducted with primers complementary to the known internal sequences.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled device camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. GENOTYPER and Sequence NAVIGATOR, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

Recently, pyrosequencing has emerged as a useful sequencing methodology. See, e.g., Ronaghi, M. et al. 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242: 84-89; Ronaghi, M. et al. 1998. A sequencing method based on real-time pyrophosphate. Science 281: 363-365; Ronaghi, M. et al. 1999. Analyses of secondary structures in DNA by pyrosequencing. Anal. Biochem. 267: 65-71; Ronaghi 2001. Genome Res. Vol. 11, Issue 1, 3-11; Nyrén The history of pyrosequencing. Methods Mol Biol. 2007;373:1-14. Pyrosequencing has the advantages of accuracy, flexibility, parallel processing, and can be easily automated. Furthermore, the technique dispenses with the need for labeled primers, labeled nucleotides, and gel-electrophoresis. In accordance with this method, polymerase catalyzes incorporation of nucleotides into a nucleic acid chain. As a result of the incorporation, pyrophosphate molecules are released and subsequently converted by sulfurylase to ATP. Light is produced in the luciferase reaction during which a luciferin molecule is oxidized. After each nucleotide addition, a washing step is performed to allow iterative addition. The nucleotides are continuously degraded by nucleotide-degrading enzyme allowing addition of subsequent nucleotide. Pyrosequencing has

been successfully applied as a platform for large-scale sequencing, including genomic and metagenomic analysis (see, e.g., The Genome Sequencer FLX™ from 454 Life Sciences/Roche).

The SOLiD™ System has also been developed for sequencing (see, e.g., Applied Biosystems. Application Fact Sheet for the SOLiD™ System. Foster City, CA). This methodology is based on sequential ligation of dye-labeled oligonucleotides to clonally amplified DNA fragments linked to magnetic beads. In this method, the DNA sequence is generated by measuring serial ligation. The ligation reaction is based on probe recognition, not sequential addition, and is therefore less prone to accumulation of errors. The nature of the chemistry virtually eliminates the possibility of spurious insertions or deletions. The ligation step and phosphatase treatment of unligated probes prevents dephasing. In addition, after seven cycles of ligation, the original primer is stripped from the template and a new primer is hybridized to begin interrogating at the n-1 position. Use of this "reset" phase allows for reduction in systemic noise and allows for longer read lengths. In addition, two base encoding is used to discriminate between measurement errors as opposed to true polymorphisms. Changes at a single position are identified as random errors and can be removed by the software in data analysis. As an analytical platform, the SOLiD™ System has applications in large-scale sequencing, digital gene expression, ChIP and methylation studies, and is particularly useful for detecting genomic variation.

In another embodiment of the invention, polynucleotides or fragments thereof which encode polypeptides may be used in recombinant DNA molecules to direct expression of the polypeptides, or fragments or variants thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express phage polypeptides. The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter amino acid-encoding sequences for a variety of reasons, including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding polypeptides may be ligated to a heterologous sequence to encode a fusion protein. For example, it may be useful to encode a chimeric sequence that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide of the invention and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding polypeptides may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the polypeptide itself may be produced using chemical methods to synthesize the amino acid sequence, or a fragment thereof. For example, polypeptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269:202-204; Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer). Various fragments of polypeptides may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

The newly synthesized polypeptide may be isolated by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic polypeptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of the polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant molecule.

In order to express biologically active polypeptides, the nucleotide sequences encoding the polypeptide or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding the polypeptide and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic

techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook, J. et al. (2001) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F. M. et al. (2007) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding the polypeptides of the invention. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant phage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. For bacteria, useful plasmids include pET, pRSET, pTrcHis2, and pBAD plasmids from Invitrogen, pET and pCDF plasmids from Novagen, and Director™ plasmids from Sigma-Aldrich. For methanogens, useful plasmids include, but are not limited to pME2001, pMV15, and pMP1. In particular, *Escherichia coli* can be used with the expression vector pET. The invention is not limited by the expression vector or host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla, CA) or pSPORT1 plasmid (Life Technologies) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the polypeptide. For example, when large quantities of polypeptide are needed, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E.*

coli cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding a polypeptide may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like.

pGEX vectors (Promega, Madison, WI) may also be used to express the polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will. In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding the polypeptides of the invention. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding a polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed polypeptide in the desired fashion. Such modifications of the sequence include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the polypeptide may also be

used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities are available from the American Type Culture Collection (ATCC; Bethesda, MD) and may be chosen to ensure the correct modification and processing of the sequence. Specific host cells include, but are not limited to, methanogen cells, such as *Methanobrevibacter* cells, in particular, *M. ruminantium*, or *M. smithii* cells. Host cells of interest include, for example, *Rhodotorula*, *Aureobasidium*, *Saccharomyces*, *Sporobolomyces*, *Pseudomonas*, *Erwinia* and *Flavobacterium*; or such other organisms as *Escherichia*, *Lactobacillus*, *Bacillus*, *Streptomyces*, and the like. Specific host cells include *Escherichia coli*, which is particularly suited for use with the present invention, *Saccharomyces cerevisiae*, *Bacillus thuringiensis*, *Bacillus subtilis*, *Streptomyces lividans*, and the like.

There are several techniques for introducing nucleic acids into eukaryotic cells cultured *in vitro*. These include chemical methods (Felgner et al., Proc. Natl. Acad. Sci., USA, 84:7413 7417 (1987); Bothwell et al., Methods for Cloning and Analysis of Eukaryotic Genes, Eds., Jones and Bartlett Publishers Inc., Boston, Mass. (1990), Ausubel et al., Short Protocols in Molecular Biology, John Wiley and Sons, New York, NY (1992); and Farhood, Annal. NY Acad. Sci., 716:23 34 (1994)), use of protoplasts (Bothwell, supra) or electrical pulses (Vatteroni et al., Mutn. Res., 291:163 169 (1993); Sabelnikov, Prog. Biophys. Mol. Biol., 62: 119 152 (1994); Bothwell et al., supra; and Ausubel et al., supra), use of attenuated viruses (Davis et al., J. Virol. 1996, 70(6), 3781 3787; Brinster et al. J. Gen. Virol. 2002, 83(Pt 2), 369 381; Moss, Dev. Biol. Stan., 82:55 63 (1994); and Bothwell et al., supra), as well as physical methods (Fynan et al., supra; Johnston et al., Meth. Cell Biol., 43(Pt A):353 365 (1994); Bothwell et al., supra; and Ausubel et al., supra).

Successful delivery of nucleic acids to animal tissue can be achieved by cationic liposomes (Watanabe et al., Mol. Reprod. Dev., 38:268 274 (1994)), direct injection of naked DNA or RNA into animal muscle tissue (Robinson et al., Vacc., 11:957 960 (1993); Hoffman et al., Vacc. 12:1529 1533; (1994); Xiang et al., Virol., 199:132 140 (1994); Webster et al., Vacc., 12:1495 1498 (1994); Davis et al., Vacc., 12:1503 1509 (1994); Davis et al., Hum. Molec. Gen., 2:1847 1851 (1993); Dalemans et al. Ann NY Acad. Sci. 1995, 772, 255 256. Conry, et al. Cancer Res. 1995, 55(7), 1397-1400), and embryos (Naito et al., Mol. Reprod. Dev., 39:153 161 (1994); and Burdon et al., Mol. Reprod. Dev., 33:436 442 (1992)), intramuscular injection of self replicating RNA

vaccines (Davis et al., J Virol 1996, 70(6), 3781 3787; Balasuriya et al. Vaccine 2002, 20(11 12), 1609 1617) or intradermal injection of DNA using "gene gun" technology (Johnston et al., supra).

A variety of protocols for detecting and measuring the expression of the polypeptides of the invention, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay can be used with monoclonal antibodies reactive to two non-interfering epitopes on the polypeptide, but a competitive binding assay can also be used. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a laboratory Manual, APS Press, St Paul, MN) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding the polypeptides, or any fragments or variants thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits Amersham Pharmacia Biotech, Promega, and US Biochemical. Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Expression vectors or host cells transformed with expression vectors may be replicated under conditions suitable for the expression and recovery of the polypeptide from culture. The culture can comprise components for *in vitro* or *in vivo* expression. *In vitro* expression components include those for rabbit reticulocyte lysates, *E. coli* lysates, and wheat germ extracts, for example, Expressway™ or RiPs systems from Invitrogen, Genelator™ systems from iNtRON Biotechnology, EcoPro™ or STP3™ systems from Novagen, TNT® Quick Coupled systems from Promega, and EasyXpress systems from

QIAGEN. The polypeptide produced from culture may be secreted or contained intracellularly depending on the sequence and/or the vector used. In particular aspects, expression vectors which encode a phage polypeptide can be designed to contain signal sequences which direct secretion of the polypeptide through a prokaryotic or eukaryotic cell membrane.

Other constructions may include an amino acid domain which will facilitate purification of the polypeptide. Such domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan (e.g., 6X-HIS) modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG® extension/affinity purification system (Immunex Corp., Seattle, WA). Useful epitope tags include 3XFLAG®, HA, VSV-G, V5, HSV, GST, GFP, MBP, GAL4, and β-galactosidase. Useful plasmids include those comprising a biotin tag (e.g., PinPoint™ plasmids from Promega), calmodulin binding protein (e.g., pCAL plasmids from Stratagene), streptavidin binding peptide (e.g., InterPlay™ plasmids from Stratagene), a c-myc or FLAG® tag (e.g., Immunoprecipitation plasmids from Sigma-Aldrich), or a histidine tag (e.g., QIAExpress plasmids from QIAGEN).

To facilitate purification, expression vectors can include a cleavable linker sequences such as those specific for Factor Xa or enterokinase (Invitrogen, San Diego, CA). For example, the vector can include one or more linkers between the purification domain and the polypeptide. One such expression vector provides for expression of a fusion protein comprising a polypeptide of the invention and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453).

Antibodies of the invention may be produced using methods which are generally known in the art, for example, for use in purification or diagnostic techniques. In particular, polypeptides or polynucleotides may be used to produce antibodies in accordance with generally known protocols. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and

fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit function) are especially preferred for use with the invention.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with a polypeptide, polynucleotide, or any fragment thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the polypeptides or fragments used to induce antibodies have an amino acid sequence comprising at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R. J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109-120). Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

In addition, techniques can be used for the production of "chimeric antibodies", e.g., the combining of antibody genes to obtain a molecule with appropriate antigen specificity and biological activity (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature

314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Those of skill in the art to which the invention relates will appreciate the terms "diabodies" and "triabodies". These are molecules which comprise a heavy chain variable domain (VH) connected to a light chain variable domain (VL) by a short peptide linker that is too short to allow pairing between the two domains on the same chain. This promotes pairing with the complementary domains of one or more other chains and encourages the formation of dimeric or trimeric molecules with two or more functional antigen binding sites. The resulting antibody molecules may be monospecific or multispecific (e.g., bispecific in the case of diabodies). Such antibody molecules may be created from two or more antibodies using methodology standard in the art to which the invention relates; for example, as described by Todorovska et al. (Design and application of diabodies, triabodies and tetrabodies for cancer targeting. J. Immunol. Methods. 2001 Feb 1;248(1-2):47-66).

Antibody fragments which contain specific binding sites may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having binding specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between a polypeptide or polynucleotide and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

The phage polypeptides described herein have the ability to target, permeabilise, and/or inhibit cells and are also useful as carrier molecules for the delivery of additional inhibitory molecules into microbial cells. The chemistry for coupling compounds to amino acids is well developed and a number of different molecule types could be linked to the polypeptides. The most common coupling methods rely on the presence of free amino (alpha-amino or Lys), sufhydryl (Cys), or carboxylic acid groups (Asp, Glu, or alpha-carboxyl). Coupling methods can be used to link the polypeptide to the cell inhibitor via the carboxy- or amino-terminal residue. In some cases, a sequence includes multiple residues that may react with the chosen chemistry. This can be used to produce multimers, comprising more than one cell inhibitor. Alternatively, the polypeptide can be shortened or chosen so that reactive residues are localized at either the amino or the carboxyl terminus of the sequence.

For example, a reporter molecule such as fluorescein can be specifically incorporated at a lysine residue (Ono et al., 1997) using *N*-α-Fmoc-*N*ε-1-(4,4-dimethyl-2,6 dioxocyclohex-1-ylidene-3-methylbutyl)-L-lysine during polypeptide synthesis. Following synthesis, 5- and 6-carboxyfluorescein succinimidyl esters can be coupled after 4,4-dimethyl-2,6 dioxocyclohex-1-ylidene is removed by treatment with hydrazine. Therefore coupling of an inhibitory molecule to the phage polypeptide can be accomplished by inclusion of a lysine residue to the polypeptide sequence, then reaction with a suitably derivatised cell inhibitor.

EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) or the carbodiimide coupling method can also be used. Carbodiimides can activate the side chain carboxylic groups of aspartic and glutamic acid as well as the carboxyl-terminal group to make them reactive sites for coupling with primary amines. The activated polypeptides are mixed with the cell inhibitor to produce the final conjugate. If the cell inhibitor is activated first, the EDC method will couple the cell inhibitor through the N-terminal alpha amine and possibly through the amine in the side-chain of Lys, if present in the sequence.

m-Maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) is a heterobifunctional reagent that can be used to link polypeptides to cell inhibitors via cysteines. The coupling takes place with the thiol group of cysteine residues. If the chosen sequence does not contain Cys it is common to place a Cys residue at the N- or C-terminus to obtain highly controlled linking of the polypeptide to the cell inhibitor. For synthesis purposes, it may

be helpful for the cysteine to be placed at the N-terminus of the polypeptide. MBS is particularly suited for use with the present invention.

Glutaraldehyde can be used as a bifunctional coupling reagent that links two compounds through their amino groups. Glutaraldehyde provides a highly flexible spacer between the polypeptide and cell inhibitor for favorable presentation. Glutaraldehyde is a very reactive compound and will react with Cys, Tyr, and His to a limited extent. The glutaraldehyde coupling method is particularly useful when a polypeptide contains only a single free amino group at its amino terminus. If the polypeptide contains more than one free amino group, large multimeric complexes can be formed.

In one aspect, the polypeptides of the invention can be fused (e.g., by in-frame cloning) or linked (e.g., by chemical coupling) to cell inhibitors such as antimicrobial agents. Included among these are antimicrobial peptides, for example, bactericidal/permeability-increasing protein, cationic antimicrobial proteins, lysozymes, lactoferrins, and cathelicidins (e.g., from neutrophils; see, e.g., Hancock and Chapple, 1999, Antimicrob. Agents Chemother.43:1317-1323; Ganz and Lehrer, 1997, Curr. Opin. Hematol. 4:53-58; Hancock et al., 1995, Adv. Microb. Physiol. 37:135-175). Antimicrobial peptides further include defensins (e.g., from epithelial cells or neutrophils) and platelet microbiocidal proteins (see, e.g., Hancock and Chapple, 1999, Antimicrob. Agents Chemother.43:1317-1323). Additional antimicrobial peptides include, but are not limited to, gramicidin S, bacitracin, polymyxin B, tachyplesin, bactenecin (e.g., cattle bactenecin), ranalexin, cecropin A, indolicidin (e.g., cattle indolicidin), and nisin (e.g., bacterial nisin).

Also included as antimicrobial agents are ionophores, which facilitate transmission of an ion, (such as sodium), across a lipid barrier such as a cell membrane. Two ionophore compounds particularly suited to this invention are the RUMENSIN™ (Eli Lilly) and Lasalocid (Hoffman LaRoche). Other ionophores include, but are not limited to, salinomycin, avoparcin, aridcin, and actaplanin. Other antimicrobial agents include penicillin, Monensin™ and azithromycin, metronidazole, streptomycin, kanamycin, and penicillin, as well as, generally, ß-lactams, aminoglycosides, macrolides, chloramphenicol, novobiocin, rifampin, and fluoroquinolones (see, e.g., Horn et al., 2003, Applied Environ. Microbiol. 69:74-83; Eckburg et al., 2003, Infection Immunity 71:591-596; Gijzen et al., 1991, Applied Environ. Microbiol. 57:1630-1634; Bonelo et al.,

1984, FEMS Microbiol. Lett. 21:341-345; Huser et al., 1982, Arch. Microbiol. 132:1-9; Hilpert et al., 1981, Zentbl. Bakteriol. Mikrobiol. Hyg. 1 Abt Orig. *C* 2:21-31).

Particularly useful inhibitors are compounds that block or interfere with methanogenesis, including bromoethanesulphonic acid, e.g., 2-bromoethanesulphonic acid (BES) or a salt thereof, for example, a sodium salt. Sodium molybdate (Mo) is an inhibitor of sulfate reduction, and can be used with bromoethanesulphonic acid. Other anti-methanogenesis compounds include, but are not limited to, nitrate, formate, methyl fluoride, chloroform, chloral hydrate, sodium sulphite, ethylene and unsaturated hydrocarbons, acetylene, fatty acids such as linoleic and cis-oleic acid, saturated fatty acids such as behenic and stearic acid, and, also lumazine (e.g., 2,4-pteridinedione). Additional compounds include 3-bromopropanesulphonate (BPS), propynoic acid, and ethyl 2-butynoate.

Further included as antimicrobial agents are lytic enzymes, including phage lysozyme, endolysin, lysozyme, lysin, phage lysin, muralysin, muramidase, and virolysin. Useful enzymes exhibit the ability to hydrolyse specific bonds in the bacterial cell wall. Particular lytic enzymes include, but are not limited to, glucosaminidases, which hydrolyse the glycosidic bonds between the amino sugars (e.g., N-acetylmuramic acid and N-acetylglucosamine) of the peptidoglycan, amidases, which cleave the N-acetylmuramoyl-L-alanine amide linkage between the glycan strand and the cross-linking peptide, and endopeptidases, which hydrolyse the interpeptide linkage (e.g., cysteine endopeptidases) and endoisopeptidases that attack pseudomurein of methanogens from the family *Methanobacteriacaea*.

The polypeptides encoded by ORF 2058 or ORF 2055, described in detail herein and below, are useful as rumen methanogen-specific lytic enzymes. The native enzymes can be prepared from freshly φmru-lysed *M. ruminantium* cells. Alternatively, ORF 2058 or ORF 2055 can be cloned in an expression vector and expressed in a heterologous host such as *Escherichia coli*. This was accomplished previously with PeiP and PeiW and the recombinant proteins were shown to be active against *Methanothermobacter* cell walls under reducing conditions (Luo et al., 2002). ORF 2058 or ORF 2055 lytic enzymes or any other lytic enzyme can be used in compositions, for example, as a feed additive for ruminants or it can be incorporated into a slow release capsule or bolus device for delivery over a longer time period within the rumen. The lytic enzymes can be used either in combination or sequentially with other methanogen inhibitor(s) to avoid

adaptation of the host methanogens and resistance to the enzymes. Random and/or targeted mutations in the enzymes can also be used to avoid adaptation. The lytic/lysogenic switch components (e.g., ORF 1981 and ORF 1983-ORF 1986) can be used in a similar manner as the lytic enzymes

Additionally, PNAs are included as antimicrobial agents. PNAs are peptide-nucleic acid hybrids in which the phosphate backbone has been replaced by an achiral and neutral backbone made from N-(2-aminoethyl)-glycine units (see, e.g., Eurekah Bioscience Collection. PNA and Oligonucleotide Inhibitors of Human Telomerase. G. Gavory and S. Balasubramanian, Landes Bioscience, 2003). The bases A, G, T, C are attached to the amino nitrogen on the backbone via methylenecarbonyl linkages (P.E. Nielsen et al., Science 1991. 254: 1497-1500; M. Egholm et al., Nature 1993. 365: 566-568). PNAs bind complementary sequences with high specificity, and higher affinity relative to analogous DNA or RNA (M. Egholm et al., supra). PNA/DNA or PNA/RNA hybrids also exhibit higher thermal stability compared to the corresponding DNA/DNA or DNA/RNA duplexes (M. Egholm et al., supra). PNAs also possess high chemical and biological stability, due to the unnatural amide backbone that is not recognized by nucleases or proteases (V. Demidov et al., Biochem Pharmacol 1994. 48: 1310-1313). Typically, PNAs are at least 5 bases in length, and include a terminal lysine. PNAs may be pegylated to further extend their lifespan (Nielsen, P. E. et al. (1993) Anticancer Drug Des. 8:53-63).

In one particular aspect, the polypeptides of the invention can be fused (e.g., by in-frame cloning) or linked (e.g., by chemical coupling) to cell inhibitors such as antibodies or fragments thereof. The antibodies or antibody fragments can be directed to microbial cells, or particularly methanogen cells, or one or more cell components. For example, cell surface proteins, e.g., extracellular receptors, can be targeted. Included are immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen.

The polypeptides of the invention find particular use in targeting a microbial cell, in particular, a methanogen cell. In certain aspects, the polypeptides can be used to associate with or bind to the cell wall or membrane, permeabilise the cell, and/or inhibit growth or replication of the cell. As such, the polypeptides can be used for transient or extended attachment to the cell, or to penetrate the cell wall or membrane and/or

accumulate in the intracellular environment. It is understood that the phage polypeptides, as well as the corresponding polynucleotides, expression vectors, host cells, and antibodies of the invention, can be used to target various microbes, for example, *Methanobrevibacter ruminantium*, which is the primary methanogen in ruminants, and *Methanobrevibacter smithii*, which is the primary methanogen in humans. To effect targeting, the microbial cell can be contacted with the phage polypeptide as isolated from one or more natural sources, or produced by expression vectors and/or host cells, or synthetic or semi-synthetic chemistry as described in detail herein. For enhanced permeabilisation, the polypeptide can be fused or linked to one or more signal sequences (predicted consensus sequence: [ML]KKKK[K]{0,1}X{0,9}[IL][IFL][IL][IL][IS][LIA]X{0,4}[LIVF][LIAV][LI][ILV][LAIV][ILFV][LIVF][SAL][ILV][GSA][AS][VAI][SA]A, see FIG. 6). See also Pérez-Bercoff, Å., Koch, J. and Bürglin, T.R. (2006) LogoBar: bar graph visualization of protein logos with gaps. Bioinformatics 22, 112-114. In particular aspects, the polypeptide is delivered to subjects as composition described in detail herein, for example, through use of a slow-release device for ruminants.

In certain embodiments, the polypeptide is fused or linked to a cell inhibitor, for example, an anti-methanogenesis compound (e.g., bromoethanesulphonic acid), an antibody or antibody fragment, lytic enzyme, peptide nucleic acid, antimicrobial peptide, or other antibiotic. The polypeptide-inhibitor is delivered to subjects as a composition to inhibit growth and/or replication of microbial cells, in particular, methanogen cells. The composition comprises, for example: a) an isolated phage, phage particle, phage genome, or alteration, fragment, variant, or derivative thereof; b) an isolated phage polypeptide, or an alteration, fragment, variant, or derivative thereof; c) an isolated polynucleotide, or an alteration, fragment, variant, or derivative thereof; d) an expression vector comprising this polynucleotide; or e) a host cell comprising this expression vector. The compositions of the invention can be specifically packaged as part of kits for targeting, permeabilising, and/or inhibiting microbial cells, especially methanogen cells, in accordance with the disclosed methods. The kits comprise at least one composition as set out herein and instructions for use in targeting or permeabilising cells, or inhibiting cell growth or replication, for methanogens or other microbes.

As an additional embodiment, the invention relates to a pharmaceutical composition in conjunction with a pharmaceutically acceptable carrier, for use with any of the methods discussed above. Such pharmaceutical compositions may comprise a phage

polypeptide, in combination with a cell inhibitor. Alternatively, the pharmaceutical compositions may comprise an expression vector or host cell as described in detail herein. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a subject alone, or in combination with other agents, drugs (e.g., antimicrobial drugs), or hormones.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA). The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the subject. Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilising agents may be added, such as the crosslinked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as

glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers. Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use. After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for

treatment of an indicated condition. For administration of a composition of the invention, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. For any compound, the therapeutically effective dose can be estimated initially either in cell assays, e.g., in microbial cells, or in particular, in methanogen cells, or in animal models, usually mice, rabbits, dogs, or pigs, or in ruminant species such as sheep, cattle, deer, and goats. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, or more, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for polynucleotides than for polypeptides. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Phage-based therapeutics are known, and methods of manufacture of such compositions are published in the art. Phage therapeutics have been described, for example, for targeting *Staphylococcus* (e.g., *S. aureus*), *Pseudomonas* (e.g., *P. aeruginosa*), *Escherichia* (e.g., *E. coli*), *Klebsiella* (e.g., *K. ozaenae*, *K. rhinoscleromatis scleromatis* and *K. pneumonia*), *Proteus*, *Salmonella*, *Shigella* (see, e.g., Carlton, R.M. (1999). Archivum Immunologiae et Therapiae Experimentalis, 47: 267-274; Liu, J. et al. (2004). Nat. Biotechnol. 22, 185-191; Projan, S. (2004). Nat. Biotechnol. 22, 167-168; Sulakvelidze, A., Alavidze, Z. and Morris, J. G. (2001). Antimicrobial Agents and Chemotherapy, 45 (3): 649-659; Weber-Dabrowska, Mulczyk, M. and Gorski, A. (2000). Archivum Immunologiae et Therapiae Experimentalis, 48: 547-551. Phage therapies have inherent advantages over traditional anti-microbials, in that phage are highly specific and don't affect the normal microflora of the human body; phage do not infect eukaryotic cells, and have no known serious side effects; phage can localize to the site of infection; and phage can replicate exponentially, so treatments require only a small dose and are generally low in cost (see, e.g., Sulakvelidze et al., supra). For current review, see Fischetti VA, Nelson D, Schuch R. Reinventing phage therapy: are the parts greater than the sum? Nat Biotechnol. 2006 Dec;24(12):1508-11.

Peptide- and polypeptide-based therapeutics have also been described, for example, for denileukin, difitox, octreotide, vapreotide, lanreotide, RC-3940 series peptides, decapeptyl, lupron, zoladex, cetrorelix (see, e.g., Lu et al., 2006, AAPS J 8:E466-472), hemocidins, staphopains (see, e.g., Dubin et al., 2005, Acta Biochemica Polonica, 52:633-638), as well as indolicidin, defensins, lantibiotics, microcidin B17, histatins, and maganin (see, e.g., Yeaman and Yount, 2003, Pharmacol Rev 55:27-55). General guidance for peptide and polypeptide therapeutics can also be found in Degim et al., 2007, Curr Pharm Des 13:99-117 and Shai et al., 2006, Curr Prot Pept Sci, 7:479-486. Recently approved peptide-based drugs include Hematide™ (synthetic peptide-based erythropoiesis-stimulating agent, Affymax, Inc.), Exenatide (synthetic exendin-4, Amylin/Eli Lilly), Natrecor (nesiritide, natriuretic peptide, Scios), Plenaxis (abarelix, Praecis Pharmaceuticals), and SecreFlo (secretin, Repligen).

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active agent or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time, and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Particularly useful for the compositions of the invention (e.g., pharmaceutical compositions) are slow release formulas or mechanisms. For example, intra-ruminal devices include, but are not limited to, Time Capsule™ Bolus range by Agri-Feeds Ltd., New Zealand, originally developed within AgResearch Ltd., New Zealand, as disclosed in WO 95/19763 and NZ 278977, and CAPTEC by Nufarm Health & Sciences, a division of Nufarm Ltd., Auckland, New Zealand, as disclosed in AU 35908178, PCT/AU81/100082, and Laby et al., 1984, *Can. J. Anim. Sci.* 64 (Suppl.), 337-8, all of which are incorporated by reference herein. As a particular example, the device can include a spring and plunger which force the composition against a hole in the end of a barrel.

As a further embodiment, the invention relates to a composition for a water supplement, e.g., drenching composition, or food supplement, e.g., ruminant feed component, for

use with any of the methods discussed above. In particular aspects, the food supplement comprises at least one vegetable material that is edible, and a peptide or polypeptide of the invention. Alternatively, the food supplement comprises at least one vegetable material that is edible, and a polypeptide or peptide, or a polynucleotide encoding a peptide or polypeptide disclosed herein, for example, as an expression vector or host cell comprising the expression vector. In particular, the composition further includes a cell inhibitor, as fused or linked to the resultant sequence. The preferred vegetable material include any one of hay, grass, grain, or meal, for example, legume hay, grass hay, corn silage, grass silage, legume silage, corn grain, oats, barley, distillers grain, brewers grain, soy bean meal, and cotton seed meal. In particular, grass silage is useful as a food composition for ruminants. The plant material can be genetically modified to contain one or more components of the invention, e.g., one or more polypeptides or peptides, polynucleotides, or vectors.

In another embodiment, antibodies which are raised to the polypeptides or polynucleotides of the invention may be used to determine the presence of microbes, especially methanogens, or in assays to monitor levels of such microbes. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above. Diagnostic assays include methods which utilize the antibody and a label to detect a polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols for measuring levels of a polypeptide or polynucleotide are known in the art (e.g., ELISA, RIA, FACS, and blots, such as Southern, Northern, Western blots), and provide a basis for determining the presence or levels of a microbe, especially a methanogen. Normal or standard levels established by combining body fluids or cell extracts taken from normal subjects, e.g., normal humans or ruminants, with the antibody under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of polypeptide or polynucleotide expressed in subject, control, and treated samples (e.g., samples from treated subjects) are compared with the standard values. Deviation between standard and subject values establishes the parameters for determining the presence or levels of the microbe.

In a particular embodiment of the invention, the polynucleotides may be used for diagnostic purposes using particular hybridization and/or amplification techniques. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in samples in which expression may be correlated with the presence or levels of a microbe. The diagnostic assay may be used to distinguish between the absence, presence, and alteration of microbe levels, and to monitor levels during therapeutic intervention.

In one aspect, hybridization with PCR probes may be used to identify nucleic acid sequences, especially genomic sequences, which encode the polypeptides of the invention. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences, alleles, or related sequences. Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the coding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:74-142, or complements, or modified sequences thereof, or from genomic sequences including promoter, enhancer elements, and introns of the naturally occurring sequence.

Means for producing specific hybridization probes for DNAs include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like. The polynucleotides may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays, or microarrays utilizing fluids or tissues from subject biopsies to detect the presence or levels of a microbe. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleic acid sequences may be useful in various assays labelled by standard methods, and added to a fluid or tissue sample from a subject under conditions suitable for hybridization and/or amplification. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the test sample is significantly altered from that of a comparable control sample, the presence of altered levels of nucleotide sequences in the sample indicates the presence or levels of the microbe. Such assays may also be used to evaluate the efficacy of a particular treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of a subject.

In order to provide a basis for the diagnosis of the presence or levels of a microbe, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, with a polynucleotide or a fragment thereof, under conditions suitable for hybridization and/or amplification. Standard levels may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from subjects treated for microbial growth. Deviation between standard and subject values is used to establish the presence or levels of the microbe.

Once the microbe is identified and a treatment protocol is initiated, hybridization and/or amplification assays may be repeated on a regular basis to evaluate whether the level of expression in the subject begins to decrease relative to that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

Particular diagnostic uses for oligonucleotides designed from the nucleic acid sequences may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'.->.3') and another with antisense orientation (3'.->.5'), employed under optimized conditions for identification of a specific nucleotide sequence or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate expression include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of disease, to diagnose disease, and to develop and monitor the activities of therapeutic agents. In one embodiment, the microarray is prepared and used according to methods known in the art such as those described in PCT application WO 95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619).

In one aspect, the oligonucleotides may be synthesized on the surface of the microarray using a chemical coupling procedure and an ink jet application apparatus, such as that described in PCT application WO 95/251116 (Baldeschweiler et al.). In another aspect, a "gridded" array analogous to a dot or slot blot (HYBRIDOT apparatus, Life Technologies) may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. In yet another aspect, an array may be produced by hand or by using available devices, materials, and machines (including multichannel pipettors or robotic instruments; Brinkmann, Westbury, N.Y.) and may include, for example, 24, 48, 96, 384, 1024, 1536, or 6144 spots or wells (e.g., as a multiwell plate), or more, or any other multiple from 2 to 1,000,000 which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarrays, polynucleotides are extracted from a biological sample. The biological samples may be obtained from any bodily fluid (blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or

other tissue preparations. To produce probes, the polynucleotides extracted from the sample are used to produce nucleic acid sequences which are complementary to the nucleic acids on the microarray. If the microarray consists of cDNAs, antisense RNAs are appropriate probes. Therefore, in one aspect, mRNA is used to produce cDNA which, in turn and in the presence of fluorescent nucleotides, is used to produce fragments or antisense RNA probes. These fluorescently labeled probes are incubated with the microarray so that the probe sequences hybridize to the cDNA oligonucleotides of the microarray. In another aspect, nucleic acid sequences used as probes can include polynucleotides, fragments, and complementary or antisense sequences produced using restriction enzymes, PCR technologies, and oligolabeling kits (Amersham Pharmacia Biotech) well known in the area of hybridization technology.

In another embodiment of the invention, the polypeptides of the invention or functional or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between the polypeptide and the agent being tested, may be measured.

One technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest as described in published PCT application WO 84/03564. In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with the polypeptide, or fragments thereof, and washed. Bound polypeptide is then detected by methods well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the polypeptide and immobilize it on a solid support.

In another technique, one may use competitive drug screening assays in which neutralizing antibodies capable of binding the polypeptide specifically compete with a test compound for binding to the polypeptide. In this manner, the antibodies can be used to detect the presence of a test compound which shares one or more antigen binding sites with the antibody.

EXAMPLES

The examples described herein are for purposes of illustrating embodiments of the invention. Other embodiments, methods, and types of analyses are within the scope of persons of ordinary skill in the molecular diagnostic arts and need not be described in detail hereon. Other embodiments within the scope of the art are considered to be part of this invention.

EXAMPLE 1: Genome size estimation

Methanobrevibacter ruminantium strain $M1^T$ (DSM1093) was grown on BY+ medium (basal medium, Joblin et al., 1990) which consists of [g/l] NaCl (1), KH_2PO_4 (0.5), $(NH_4)_2SO_4$ (0.25), $CaCL_2.2H_2O$ (0.13), $MgSO_4.7H_2O$ (0.2), K_2HPO_4 (1), clarified rumen fluid (300 ml), dH_2O (360 ml), $NaHCO_3$ (5), resazurin (0.2 ml), L-cysteine-HCl (0.5), yeast extract (2), and Balch's trace elements solution (10 ml) (added trace elements; Balch et al., 1979) which consists of (g/l) nitrilotriacetic acid (1.5), $MgSO_4.7H_2O$ (3), $MnSO_4.H_2O$ (0.5), NaCl (1), $FeSO_4.7H_2O$ (0.1), $CoCl_2.6H_2O$ (0.1), $CaCl_2$ (0.1), $ZnSO_4.7H_2O$ (0.1), $CuSO_4.5H_2O$ (0.01), $AlK(SO_4)_2.12H_2O$ (0.01), H_3BO_3 (0.01), $Na_2MoO_4.2H_2O$ (0.01), $NiSO_4.6H_20$ (0.03), Na_2SeO_3 (0.02), and $Na_2Wo_4.2H_2O$ (0.02). Genomic DNA was extracted by freezing cell pellets in liquid N_2 and grinding using a pre-chilled, sterilised mortar and pestle. Cell homogenates were imbedded in agarose plugs and subsequent manipulations were carried out in the plugs to reduce the physical shearing of genomic DNA. Digests were performed with restriction endonucleases and DNA fragments were separated using pulsed-field gel electrophoresis (PFGE).

EXAMPLE 2: DNA cloning and sequencing

The DNA of the *M. ruminantium* genome was sequenced by Agencourt Biosciences Corporation (Massachusetts, USA) using a random shotgun cloning approach (Fleischmann et al., 1995) and by Macrogen Corporation (Rockville, MD, USA) using pyrosequencing. Briefly, libraries of *M. ruminantium* DNA were constructed in *Escherichia coli* by random physical disruption of genomic DNA and separation of fragments by gel electrophoresis. Large fragments in the 40 Kb range were retrieved from the gel and used to generate a large insert fosmid library. DNA fragments in the 2 to 4 Kb range were recovered and used to generate a small insert plasmid library. Clones resulting from both large and small insert libraries were grown, and their fosmid or plasmid DNA was recovered and sequenced using high throughput sequencing technology. A sufficient number of clones were sequenced to give a theoretical 8 fold

coverage of the *M. ruminantium* genome. Pyrosequencing was performed on randomly sheared genomic DNA fragments to give a final theoretical 10 fold coverage.

EXAMPLE 3: Sequence assembly and prophage annotation

DNA sequences were aligned to find sequence overlaps and assembled into contiguous (contig) sequences using Paracel Genome Assembler (Paracel Inc, CA, USA) and the Staden package (Staden *et al.,* 1998) in combination with sequence from both standard and inverse PCRs. Contigs were analysed using the open reading frame (ORF) finder GLIMMER (Gene Locator Interpolated Markov Model ER, Delcher et al., 1999) and each ORF was analysed by gapped BLASTP (Basic Local Alignment Search Tool (Altschul et al., 1997) against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide and protein databases. The contigs from the 8 fold draft phase sequence were joined at random by artificial linking of sequences to generate a "pseudomolecule" and submitted to The Institute for Genomic Research (TIGR, DC, USA) for autoannotation. The contigs assembled from the 10 fold pyrosequencing were reanalysed using GLIMMER and ORFs were autoannotated using GAMOLA (Global Annotation of Multiplexed On-site Blasted DNA sequences; Altermann and Klaenhammer, 2003). ORFs were categorised by function using the clusters of orthologous proteins (COG) database (threshold 1e-02) (hypertext transfer protocol://world wide web.pnas.org/cgi/content/full/102/11/3906; Tatusov et al., 2001).

Protein motifs were determined by HMMER (hypertext transfer protocol://hmmer.wustl.edu) using PFAM HMM and TIGRFAM libraries, with global and local alignment (hypertext transfer protocol://pfam.wustl.edu) and standard and fragment-mode TIGRFAM HMMs models (hypertext transfer protocol://world wide web.tigr.org/TIGRFAMs) respectively (threshold 1e-02). tRNAs were identified by using TRNASCAN-SE (Lowe and Eddy, 1997) and nucleotide repeats were identified using the KODON software package (Applied Maths, Austin, TX, USA) and REPUTER (Kurtz and Schleiermacher, 1999). Automated annotations were subsequently verified manually. Genome atlas visualizations were constructed using GENEWIZ (Jensen et al., 1999) and underlying data structures were generated by customised in-house developed algorithms. Pathway reconstructions from the predicted *M. ruminantium* ORFeome were carried out in conjunction with the KEGG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa et al., 2004) on-line database using in-house developed software (PathwayVoyager; Altermann and Klaenhammer, 2005).

EXAMPLE 4: Sequencing results and analysis

Size estimation of the *M. ruminantium* genome by restriction enzyme digestion of genomic DNA and sizing of fragments via PFGE, indicated a single chromosome of approximately 2.5-2.9 Mb. Initial sequencing of large and small insert clones (6 fold draft coverage) and assembly of the sequence into contigs indicated that a 40 Kb region of the genome was highly over-represented (>20 fold), particularly within the small insert library. This was possibly due to a high copy number plasmid (although no extrachromosomal DNAs had been identified) or a lysogenic bacteriophage that had replicated during the growth of the culture used for DNA extraction. Because of this large sequence bias, additional sequencing was carried out (2 fold theoretical genome coverage) for only small insert clones yielding a final 8 fold coverage from Sanger sequencing. The 8 fold draft phase sequence was assembled into 756 contigs which were linked via 105 scaffolds. Further pyrosequencing was carried out to an additional ~10 fold coverage and incorporation of these sequences into the assembly resulted in the contig number dropping to 27. Subsequent gap closure using inverse and long range PCR techniques reduced the contig number to 14, with one misassembly remaining.

During the high-throughput sequencing phase, a bias was observed in the sequence coverage towards a region (~50Kb) of significantly higher G+C content immediately adjacent to a low G+C region (~12Kb). Analysis of the genome sequence via GAMOLA and GeneWiz led to the discovery of a prominent high-GC region located immediately adjacent to a large low-GC spike. Detailed analyses of the high G+C region revealed the presence of gene-products with similarities to a phage-related integrase, the large subunit of the phage terminase, a phage portal protein, a phage capsid protein, and a predicted peptidase acting as phage lysin (FIG. 3). These gene products were used as anchor points for the overall structure of the predicted *M. ruminantium* prophage, designated φmru. Based on analyses of DNA secondary structures, the likely phage integration sites *att*L and *att*R were identified (FIG. 1A). Phage integration at the *att* site appears to have disrupted a putative membrane protein encoded by ORFs 1980 and 2069, and this gene may harbour the original integration site for the φmru phage genome, *attB*.

The general structure (FIG. 1B) and DNA sequence (FIG. 4A) of φmru were determined based on commonly recognized modular structure of phage genomes combined with similarities to sequence and functional databases. See, e.g., Altermann E, Klein JR,

Henrich B. Primary structure and features of the genome of the *Lactobacillus gasseri* temperate bacteriophage (phi)adh. Gene. 1999 Aug 20;236(2):333-46; Desiere F, Lucchini S, Canchaya C, Ventura M, Brüssow H. Antonie Van Leeuwenhoek. Comparative genomics of phages and prophages in lactic acid bacteria. 2002 Aug;82(1-4):73-91. The predicted φmru phage ORFeome was successfully classified into modules encoding phage integration, DNA replication and packaging, phage structural proteins, and a lysis cassette, and approximately 40% of the phage ORFs were functionally characterised. A terminator-like structure in a large non-coding region (244 bp), flanked by a large number of direct and indirect repeats and determined within the DNA replication module was characterised as a putative origin of DNA replication. Several genes within the phage genome sequence were predicted on the antisense strand and these coincided with low-GC regions. It is to be determined if these genes inactivate phage function or indicate misassembley within the phage genome.

The low-GC region between the predicted phage lysin and *att*R, was found to harbour a DNA sulphur modification system, dnd (degradation during electrophoresis), including a type II restriction m6 adenine DNA methyltransferase and a transcriptional regulator likely to be specific for the dnd system. Furthermore, non-coding RNA structures were identified both within and flanking the phage genome. Within the predicted DNA replication module, an rbcL was identified. rbcL represents a 5' UTR RNA stabilising element from *Chlamydomonas reinhardtii*. The family is thought to be involved in the stabilisation of the rbcL gene which codes for large subunit of ribulose-1,5-bisphosphate carboxylase. Mutations in this family can lead to a 50-fold acceleration in transcript degradation.

Flanking the phage genome, three group I intron structures were identified. Group I catalytic introns are large self- splicing ribozymes. They catalyse their own excision from mRNA, tRNA and rRNA precursors in a wide range of organisms. The core secondary structure consists of 9 paired regions (P1-P9). These fold to essentially two domains - the P4-P6 domain (formed from the stacking of P5, P4, P6 and P6a helices) and the P3-P9 domain (formed from the P8, P3, P7 and P9 helices). The secondary structure mark-up for this family represents only this conserved core. Group I catalytic introns often have long ORFs inserted in loop regions. These non-coding RNA structures are located in the non-coding regions between upstream of ORF 1980 (SEQ ID NO:74), downstream of ORF 2065 (SEQ ID NO:141) and attR and upstream of ORF 2069 (SEQ ID NO:142).

EXAMPLE 5A: Phage genes

The discovery of a prophage sequence within the *M. ruminantium* genome sequence was unexpected. There have been no previous reports of *Methanobrevibacter ruminantium* strain M1 (DSM 1093) being susceptible to either lytic or lysogenic phage, although there have been reports of phage being identified for other Methanobrevibacter species (Baresi and Bertani, 1984; Knox and Harris, 1986). The sequence of the φmru prophage is significantly higher in G+C content than the surrounding *M. ruminantium* genome suggesting that it originated from another organism. The observed levels of homology do not suggest an obvious host from which it originated, and indicates that φmru is unlike any other phage encountered to date.

The φmru DNA sequence is inserted within a predicted *M. ruminantium* membrane protein and is flanked by DNA sequences with secondary structures consistent with *attL* and *attR* sites. Despite the lack of strong homology to other known proteins, all of the functional module characteristics of a phage could be identified within the φmru sequence. An interesting feature of the sequence is a low G+C region at the 3' end which shows homology to proteins involved in a DNA sulphur modification system (dnd) system. These genes are located upstream of the *attR* attachment site of φmru and were therefore likely brought into the *M. ruminantium* genome during phage integration. The region encodes several dnd associated ORFs (dnd 1, 2, and 3), and a Type II methylase subunit along with a putative transcriptional regulator.

The dnd phenotype sensitises its DNA to degradation during electrophoresis. Analyses of respective dnd ORF functions suggested an incorporation of sulphur or a sulphur-containing substance into the hosts' genome. The Dnd phenotype was also discovered to exist in DNA of widespread bacterial species of variable origin and diverse habitat. Similarly organized gene clusters were found in several bacterial genomes representing different genera and in eDNA of marine organisms, suggesting such modification as a widespread phenomenon. A coincidence between the Dnd phenotype and DNA modification by sulphur was demonstrated to occur in several representative bacterial genomes by the *in vivo* (35)S-labelling experiments (Zhou X, He X, Liang J, Li A, Xu T, Kieser T, Helmann JD, Deng Z. A novel DNA modification by sulphur. Mol Microbiol. 2005 Sep;57(5):1428-38).

Type II R\M systems are the simplest and the most prevalent. Instead of working as a complex, the methytransferase and endonuclease are encoded as two separate proteins and act independently. There is no specificity protein. Both proteins recognize the same recognition site, and therefore compete for activity. The methyltransferase acts as a monomer, methylating the duplex one strand at a time. The endonuclease acts as a homodimer, which facilitates the cleavage of both strands. Cleavage occurs at a defined position close to or within the recognition sequence. At this point it is unclear how the predicted functionality acts together with the dnd system. Yet, it is clear that the phage has utility as a gene delivery vehicle. In particular, the lysogenic conversion region can be used as the locus of gene replacement.

It is likely that the dnd system was transported to *M. ruminantium* by the phage. As such, the role of the φmru dnd system in protecting or modifying *M. ruminantium* or foreign DNA is unknown. Another interesting feature of the φmru sequence is the number of ORFs encoded on the antisense strand. These ORFs correspond with low GC regions and have weak BLAST matches to proteins from a variety of organisms. This could suggest that these genes have been accumulated within the φmru genome since its integration into *M. ruminantium*. It is not clear if these ORFs represent an ongoing accumulation of insertions that may eventually lead to phage inactivation and domestication or if φmru is fully active.

One φmru gene of particular interest to methane mitigation is ORF2058 located in the lysis cassette. ORF 2058 is annotated as a peptidase and has a Protein Family (Pfam) match (Score:-13.7, E value:0.00054) to Peptidase C39 family proteins. These proteins are cysteine peptidases and are part of the larger clan of CA peptidases as defined by the MEROPS peptidase database (Rawlings et al., 2006). The C39 peptidase family are usually associated with ABC transporters and function as maturation proteases during the export and processing of bacteriocins. The CA peptidase clan also includes the viral cysteine endopeptidases such as the C71 archaeal phage endoisopeptidases that cleave the crosslinking peptides of the archaeal cell wall. The cell walls of methanogenic archaea belonging to the *Methanobacteriales* family contain parallel chains of pseudomurein, a polymer of *N*-acetyl-L-talosaminurinic acid crosslinked by a peptide. The C71 pseudomurein endoisopeptidases are able to cleave the cell wall peptide crosslinks of archaea and lyse the cells.

Based on the location and synteny with the pseudomurein endoisopeptidase from *Methanothermobacter marburgensis* phage ΨM2 (FIG. 3), ORF 2058 may have a role as a methanogen lysin gene which encodes the lytic enzyme involved in cell lysis prior to release of phage progeny. Alignment of ORF 2058 with PeiP from *M. marburgensis* and PeiW from *M. wolfeii* (FIG. 5) shows low overall homology between the proteins. However there is conservation of the histidine and aspartic acid residues involved in the endoisopeptidase catalytic triad and a cysteine residue in ORF 2058 is positioned near the conserved cysteine of PeiP and PeiW which makes up the third conserved site of the catalytic triad (Makarova *et al.*, 1999, Luo *et al.*, 2002). Furthermore, the Gly-His-Tyr motif surrounding the catalytic His residue in PeiP and PeiW is also found in ORF 2058. These observations indicate that ORF 2058 is a φmru lysin gene which functions to lyse *M. ruminantium* cells during the phage lytic cycle. The differences observed between ORF 2058 and PeiP and PeiW may reflect different archaeal cell wall peptide crosslinks and therefore peptidase substrate specificity.

EXAMPLE 5B: Phage induction

Methanobrevibacter ruminantium strain $M1^T$ (DSM1093) was grown on BY+ medium (basal medium, Joblin et al., 1990) which consists of [g/l] NaCl (1), KH_2PO_4 (0.5), $(NH_4)_2SO_4$ (0.25), $CaCL_2.2H_2O$ (0.13), $MgSO_4.7H_2O$ (0.2), K_2HPO_4 (1), clarified rumen fluid (300 ml), dH_2O (360 ml), $NaHCO_3$ (5), resazurin (0.2 ml), L-cysteine-HCl (0.5), yeast extract (2), and Balch's trace elements solution (10 ml) (added trace elements; Balch et al., 1979) which consists of (g/l) nitrilotriacetic acid (1.5), $MgSO_4.7H_2O$ (3), $MnSO_4.H_2O$ (0.5), NaCl (1), $FeSO_4.7H_2O$ (0.1), $CoCl_2.6H_2O$ (0.1), $CaCl_2$ (0.1), $ZnSO_4.7H_2O$ (0.1), $CuSO_4.5H_2O$ (0.01), $AlK(SO_4)_2.12H_2O$ (0.01), H_3BO_3 (0.01), $Na_2MoO_4.2H_2O$ (0.01), $NiSO_4.6H_20$ (0.03), Na_2SeO_3 (0.02), and $Na_2Wo_4.2H_2O$ (0.02).

At optical densities (OD), measured at a wavelength of 600 (OD_{600}), between 0.10 and 0.14, *M. ruminantium* was challenged with 1 ml and 2 ml of sterile air (~160 to 320 µl oxygen), respectively (Figure 1C) and 2 µg/ml MitomycinC (Figure 1D). Typical lysis curves could be observed for both challenges, with latent times of ~90 min for air challenge. Initial results for MitomycinC challenge indicate a very short latent period. To verify the excision of the phage from the host genome, 2 oligonucleotides were designed, facing both phage attachment sites, respectively (R1F: caaagagagattaaagaagcagacg; SEQ ID NO:146 and L2R agtagtgttggaatcagtgaaaagg; SEQ ID NO:147). This primer pair only produces an amplicon if the phage genome recircularises upon excision.

Figure 1E depicts the initial excision experiments when *M. ruminantium* was challenged with air. Upon induction, a clear and unambiguous amplicon of the expected size was found, indicating successful excision and recircularisation. A similar, albeit weaker band was also found in uninduced *M. ruminantium* cells, indicating that φmru has the ability to spontaneously excise during normal, unchallenged growth.

EXAMPLE 5C: Lytic Enzyme Bioassays

The polypeptide encoded by ORF 2058 is useful as a rumen methanogen-specific lytic enzyme and has been sub-cloned in an *E. coli* expression vector for production of recombinant protein. ORF 2058 was amplified by PCR using the primers Mbbrum11for22 (1122For, cac cat ggt tag att cag cag aga c; SEQ ID NO:148) and Mbbrum11rev22 (1122Rev, tca tgc agg aca gac aac ata gta g; SEQ ID NO:149) in 150 µL reaction volume containing: 121.5 ng *M. ruminantium* strain M1 genomic DNA; 0.2 µM 1122For and 1122Rev primers;15 µL Accuprime Pfx buffer (with dNTPs, InVitrogen); 2.4 µL Accuprime Pfx (InVitrogen).PCR conditions were 95°C for 2 min initial denaturation followed by 35 cycles of 95°C for 15 seconds, 55°C for 30 seconds and 68°C for 40 seconds. No final extension was used. The PCR product was purified and quantified using a Nanodrop (Thermo Scientific, GA, USA).

ORF 2058 cloning: The PCR-amplified ORF 2058 was cloned into either pET 100 or pET 151-D Topo vectors (InVitrogen) according to the manufacturer's recommendations, and transformed into chemically competent TOP 10 cells (InVitrogen). Transformants were analysed by colony PCR, and plasmid DNA purified and sequenced. Clones with DNA sequences matching that of ORF 2058 were selected.

ORF 2058 expression: Plasmid DNA from clones containing verified ORF 2058 inserts were transformed via electroporation into electro-competent BL21* or Rosetta 2 cells. The best growth conditions for expression of soluble ORF 2058 protein was found to be in LB media, with induction being carried out between 0.48-0.6 Absorbance 600 nm using 0.5 mM IPTG and continuing growth for approximately six hours at 30°C. Cells were then harvested by centrifugation and frozen at -20°C.

Cell lysis: The cell pellet was thawed and resuspended in the following buffer (pH 7.5): 300 mM NaCl, 2 mM DTT, 10 mM imidazole, 20 mM Tris, 20% glycerol, 1% Triton-X, 5 mM $CaCl_2$, and 10 mM $MgCl_2$. Lysozyme was added to 1 mg/ml final concentration,

followed by incubation on ice with gentle agitation for 30 min. DNase I and RNase I were each added to 5 µg/ml final concentration followed by incubation on ice with gentle agitation for 30 min. The cell lysate was centrifuged at 12,000 rpm for 15 min and the crude lysate was filtered through a 0.8 µm filter.

Nickel affinity chromatography: The filtered supernatant from the cell lysis procedure was applied to an 80 mL nickel affinity column and eluted using a 20 mM to 250 mM imidazole gradient in the following buffer (pH 7.5): 300 mM NaCl, 2 mM DTT, 20 mM Tris, and 20% glycerol. Fractions eluted from the column containing the expressed ORF 2058 protein were concentrated using a Millipore ultra filtration cell with a 10,000 kDa molecular weight cut-off membrane. The ORF 2058 construct in pET100 expressed in *E. coli* BL21* cells was eluted from the nickel column by the following elution buffer, pH 8.2 (20 mM Tris, 250 mM imidazole, 300 mM NaCl, 10 mM b-mercaptoethanol, 10% glycerol), and the enzyme was stored in a buffer in which additional glycerol and dithiothreitol were added to achieve a final concentration of 40% glycerol, 1 mM dithiothreitol, pH 8.2)

Desalting: Desalting of the concentrated protein expressed from the pET 151 construct in Rosetta 2 cells was performed using a 250 mL BioGel P6 DG (BioRad, CA, USA) column with the following buffer (pH 7.0):20 mM MOPS, 1 mM DTT, 300 mM NaCl, and 20% glycerol. Fractions from the column were concentrated as described above and the final sample was filtered and snap-frozen in liquid nitrogen before being stored at -20°C.

Lysis of resting *M. ruminantium* cells: Five ml cultures of *M. ruminantium* M1 (DSM 1093) were grown in BY+ medium in Hungate tubes to late log phase and cells were collected by centrifugation of the Hungate tubes at 5,000 x *g* at room temperature for 30 minutes. The tubes were moved into an anaerobic chamber (95% CO_2- 5% H_2 atmosphere, Coy Laboratory Products, MI, USA) where the supernatant was discarded and the cells from 10 ml culture were resuspended in 1 ml MOPS buffer pH 6.8 (50 mM MOPS, 5 mM $CaCl_2$, 1 mM dithiothreitol). The cell suspension was adjusted to an OD (600 nm) of ~0.12 by dilution with additional MOPS buffer.

The standardised cell suspension (50 µl) was dispensed into a microtitre plate and varying concentrations of ORF 2058 lytic enzyme (prepared from the pET 100 construct) were added and the total volume of the reaction was made up to 250 ml with buffer. The cell and protein mixtures were incubated at 37°C and OD readings were

recorded. The effects of the enzyme additions (µg enzyme added per assay) on resting *M. ruminantium* cells are shown in FIG. 7. The enzyme additions decreased the OD 600 nm readings of the suspended cells in a dose-dependant manner compared to the control cells without added enzyme. This indicates that the ORF 2058 lytic enzyme is able to attack and lyse resting cells of *M. ruminantium* under anaerobic conditions.

Lysis of growing *M. ruminantium* cells: *M. ruminantium* was grown in RM02 medium. RM02 medium was composed of the following ingredients (g/L): KH_2PO_4 (1.4), $(NH_4)_2SO_4$ (0.6), KCl (1.5), trace element solution SL10 (1 ml), selenite/tungstate solution, (1 ml), 0.1% (w/v) resazurin solution (4 drops). The components were mixed and boiled under O_2-free 100% CO_2 and cooled in an ice bath while bubbling with 100% CO_2. After cooling $NaHCO_3$ (4.2 g) and L-cysteine·HCl·H_2O (0.5 g) were added and 9.5 ml of the medium was dispensed into Hungate tubes while gassing the tubes with 100% CO_2. The tubes were autoclaved and stored in the dark for 24 h before using. Prior to inoculation, NoSubRFV (0.5 ml per tube, containing substrates, yeast extract, vitamins) was added. After inoculation tubes were gassed with 80% CO_2/20% H_2 to 25 lb/in^2. *M. ruminantium* was grown to mid-log (OD 600 nm ~0.1) at which point ORF 2058 lytic enzyme (prepared from the pET 151 D Topo clone) was added to cultures at varying concentrations. Incubation of cultures continued and OD readings were recorded. The effect of the enzyme additions on *M. ruminantium* growth and methane formation (% methane production relative to the no-addition control after 217 hours growth are indicated in brackets) are shown in FIG. 8. The results show that the ORF 2058 lytic enzyme dramatically affected the growth of *M. ruminantium* in a dose-dependant manner, decreasing the OD 600 nm of growing cultures within 2 hours of addition. The two highest levels of enzyme addition also reduced methane formation to an extent similar to that of chloroform addition (100 µl /10 ml culture addition).

EXAMPLE 6: Overview

An unexpected discovery from the sequencing of the *M ruminantium* genome was the presence of a prophage sequence. Analysis of the genome sequence identified a region of unusually high GC content which contained a number of phage-related genes. The overall structure of the predicted prophage was identified by further bioinformatic analyses and designated as φmru. Approximately 40% of the phage genes were assigned to discrete functional groups including phage integration, DNA replication and packaging, phage structural proteins and lysis. DNA sequences flanking the phage genome were found to represent potential sites for phage integration *(attL* and *attR).*

The phage appears to have inserted itself into a *M. ruminantium* putative membrane protein which likely harbours the original methanogen integration site for the φmru phage genome, *attB*. Furthermore, a terminator-like structure found within the DNA replication module is thought to represent an origin of phage DNA replication. A low-GC region at the 3' end of the phage genome harbours what appears to be a DNA modification system by sulphur, including a gene that is likely to control the expression of the dnd system. These genes were probably carried into the M. *ruminantium* genome during phage integration and their role with respect to modifying phage, host or foreign DNA remains to be elucidated. The retention of the dnd system by *M ruminantium* suggests it has imparted a benefit to the host. However the role of the φmru dnd system in modifying *M ruminantium* or foreign DNA is still under investigation.

Another interesting feature of the φmru sequence is the number of genes encoded on the antisense strand which correspond with low GC regions and have weak matches to proteins from a variety of organisms. This suggests that these genes have accumulated within the φmru genome since its integration into *M. ruminantium* and it may be that these genes represent an ongoing build up of insertions that might eventually lead to phage inactivation and phage domestication. The high GC content of the φmru phage sequence compared to the *M. ruminantium* genome suggests that it originated from another organism. However, the previous host is not obvious as the φmru proteins appear somewhat unique by comparison to other phage encountered to date.

The φmru genes of notable interest in regard to methane mitigation are those located within the lysis cassette. One gene in particular encodes a protein with similarity to family C39 peptidases. This peptidase family includes, among others, viral cysteine endopeptidases such as the C71 archaeal phage endoisopeptidases that cleave the crosslinking peptides of pseudomurein which makes up *Methanobrevibacter* cell walls. Based on gene location within the phage genome and synteny with pseudomurein endoisopeptidases from other non-rumen methanogen phage genomes, this gene may have a role as a lysin gene encoding the lytic enzyme involved in cell lysis prior to release of phage progeny. This gene and its encoded enzyme are of obvious interest as possible control mechanism for *M. ruminantium* and other rumen methanogens with similar cell walls.

Ruminant phage and their enzymes that are involved in lysing host cells represent significant opportunities for controlling both methanogen populations and other

community members (bacteria, protozoa and fungi) in the rumen. In addition, it is possible to identify key host enzyme targets that are susceptible to inhibition by phage proteins through understanding the life cycles of phage. The inventors have surveyed the composition of rumen phage in cows, sheep and deer and shown them to display temporal variation in numbers and type. New Zealand methanogen isolates that are affected by phage have also been identified. Pure cultures of methanogens have been used to evaluate phage lytic enzymes, and culture-based and PCR-based techniques have been developed to screen for novel phage. Purified phage from rumen samples have been shown to be amenable to random DNA sequence analysis which enables phage enzymes to be discovered.

There are several advantages to the use of phage or their enzymes in mitigation techniques for lowering methane emissions. Phage are natural members of the rumen microbial community and, thus, would not be viewed as antibiotic treatment (and could more easily overcome any regulatory constraints). Phage are usually specific for a narrow range of hosts potentially enabling the selected targeting of methanogens. Phage therapy is now recognised as a treatment for antibiotic resistant organisms and generally regarded as safe. Once produced, phage are usually relatively stable. Introduction of methanogens strains into the rumen that are susceptible to phage could have long-term beneficial effects, particularly if inoculation occurs at an early age (e.g., in young lambs and calves). Certain methanogens are known to either contain phage genomes, be susceptible to lytic phage, or undergo autolysis (suggestive of lytic enzymes) including *Methanobrevibacter smithii* (strain PS), *Methanobacterium bryantii* and *Methanobrevibacter* strain MF-1. One notable example of phage being used to inhibit agriculturally problematic organisms is the use of phage to target *Escherichia coll.* 0157:H7.

Methanobrevibacter ruminantium was chosen for genome sequencing because of its prevalence in the rumen under a variety of dietary conditions (based on cultivation and molecular detection data), the availability of cultures, its amenity to routine growth in the laboratory, and the relatively large amount of previous studies and background literature available for this organism. The present invention provides important data regarding the *M. ruminantium* genome, and constructs a detailed picture of the phage within the rumen. The φmru prophage sequence provides specific reagents for inhibition of *M. ruminantium* and for future genetic manipulations to assist in determining gene function.

The phage can be used to block conserved functions/components among methanogens to prevent or reduce methane formation in the rumen.

References

Altermann E, Klaenhammer TR (2005) PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* **6**:60-66.

Altermann, E Klaenhammer TR (2003) GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *OMICS: A journal of integrative biology* **7**, 161–169.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25,** 3389–3402.

Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. *Microbiological Reviews* **43**, 260-296.

Baresi, L. and Bertani, G. 1984. Isolation of a bacteriophage for a methanogenic bacterium. In *Abstracts of the Annual Meeting of the American Society for Microbiology*. Washington DC: American Society for Microbiology, p. 133.

Bickle, T. A. and D. H. Kruger. 1993. Biology of DNA restriction. Microbiol. Rev. 57:434-450.

Bult CJ, et al. (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* **273**, 1058-1073.

Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In 'Recent Advances in Carbohydrate Bioengineering' (Eds HJ Gilbert, G Davies, B Henrissat and B Svensson) pp. 3-12 (The Royal Society of Chemistry, Cambridge) (Carbohydrate Active Enzymes database, http://www.cazy.org/).

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* **27**, 4636–4641.

Fleischmann, RD, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Fricke WF, Seedorf H, Henne A, Kruer M, Liesegang H, Hedderich R, Gottschalk G, Thauer RK (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. *Journal of Bacteriology* **188**, 642-658.

Godde JS, Bickerton A (2006) The repetitive DNAe called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution* **62,** 718-729.

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Computational Biology* **1**:474-483

Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43,** 1565-1575.

Jansen R, van Embden JD, Gaastra W, Schouls LM (2002) Identification of a novel family of sequence repeats among prokaryotes. *OMICS: A journal of integrative biology* **6**, 23-33.

Jensen LJ, Friis C, Ussery DW (1999) Three views of microbial genomes. *Research in Microbiology* **150,** 773–777.

Joblin KN, Naylor GE, Williams AG (1990) Effect of *Methanobrevibacter smithii* on xylanolytic activity of anaerobic ruminal fungi. *Applied and Environmental Microbiology* **56**, 2287-2295.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32,** D277–D280.

Kiener, A., Konig, H., Winter, J. and Leisinger, T. 1987. Purification and use of *Methanobacterium wolfeii* pseudomurein endopeptidase for lysis of *Methanobacterium thermoautotrophicum.* J. Bacteriol. 169, 1010-1016.

Knox, M. R. and Harris, J. E. 1986. Isolation and characterisation of a bacteriophage of *Methanobrevibacter smithii.* In *Abstracts of the XIV International Congress on Microbiology.* Manchester: International Union of Microbiological Societies.

Kurtz S, Schleiermacher C (1999) REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* **15,** 426–427.

Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25,** 955–964.

Loenen, W. and N. Murray. 1986. Modification enhancement by restriction alleviation protein (Ra1) of bacteriophage lambda. J. Mol. Biol. 190:11-22.

Lucchini, S., F. Desiere, and H. Brussow. 1999. Comparative genomics of *Streptococcus thermophilus* phage species supports a modular evolution theory. J. Virol. 73:8647-8656.

Luo, Y. N., Pfister, P., Leisinger, T. and Wasserfallen, A. 2002. Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family

of proteases produced in *Methanothermobacter* strains. . FEMS Microbiol. Lett. 208, 47-51.

Makarova, K. S., Aravind, L. and Koonin, E. V. 1999. A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. Protein Sci. 8, 1714-1719.

Makarova KS, Grishin NV, Shabalina, SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**:7-32.

New Zealand Statistics 2005 (www.stats.govt.nz)

New Zealand's Greenhouse Gas Inventory 1990 – 2004. The National Inventory Report and Common Reporting Format. (2006) Ministry for the Environment. http://www.mfe.govt.nz/publications/climate/nir-apr06/nir-apr06.pdf.

Reeve JN, Nolling J Morgan RM, Smith DR (1997) Methanogenesis: genes, genomes and who's on first? *Journal of Bacteriology* **179**, 5975-5986.

Rawlings, N. D., Morton, F. R. and Barrett, A. J. 2006. *MEROPS*: the peptidase database. Nucleic Acids Res. 34, D270-D272.

Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK, Gordon JI (2007) Genomic adaptations of *Methanobrevibacter smithii* to the human gut. *Proceedings of the National Academy of Sciences USA* **104**, 10643-10648.

Smith DR, et al. (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* □H: Functional analysis and comparative genomics. *Journal of Bacteriology* **179,** 7135–7155.

Smith PH, Hungate RE (1958) Isolation and characterization of *Methanobacterium ruminantium* n. sp. *Journal of Bacteriology* **75,** 713-718.

Staden R, Beal KF, Bonfield JK (1998) The Staden Package. *Methods in Molecular Biology: Bioinformatics Methods and Protocols* **132**, 115-130.

Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes *Nucleic Acids Research* **29,** 22–28.

All publications and patents mentioned in the above specification are herein incorporated by reference.

Where the foregoing description reference has been made to integers having known equivalents thereof, those equivalents are herein incorporated as if individually set forth. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. It is appreciated that further modifications may be made to the invention as described herein without departing from the spirit and scope of the invention.

WHAT WE CLAIM IS:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2-5.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:62.

3. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:63 and 72.

4. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:64-68.

5. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, 6-61, and 69.

6. An isolated polypeptide which shares 90% identity with an amino acid sequence selected from the group consisting of SEQ ID NO: 2-5.

7. An isolated polypeptide which shares 90% identity with an amino acid sequence selected from the group consisting of SEQ ID NO:62.

8. An isolated polypeptide which shares 90% identity with an amino acid sequence selected from the group consisting of SEQ ID NO:63 and 72.

9. An isolated polypeptide which shares 90% identity with an amino acid sequence selected from the group consisting of SEQ ID NO: 64-68.

10. An isolated polypeptide which shares 90% identity with an amino acid sequence selected from the group consisting of SEQ ID NO:1, 6-61, and 69.

11. An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of SEQ ID NO:2-5.

12. An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of SEQ ID NO:62.

13. An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of SEQ ID NO:63 and 72.

14. An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of SEQ ID NO:64-68.

15. An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of SEQ ID NO:1, 6-61, and 69.

16. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:75-78.

17. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:135.

18. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:136.

19. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:137-141.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:74, 79-134, and 142.

21. A vector encoding the polypeptide of any one of claims 1 to 10.

22. A vector comprising the polynucleotide of any one of claims 11 to 20.

23. A host cell comprising the vector of claim 21 or claim 22.

24. A host cell which is genetically modified to encode the polypeptide of any one of claims 1 to 10.

25. A host cell which is genetically modified to include the polynucleotide of any one of claims 11 to 20.

26. The host cell of any one of claims 23 to 25, which is prokaryotic.

27. The host cell of claim 26, which is *Escherichia coli.*

28. The host cell of any one of claims 23 to 25, which is a methanogen.

29. The host cell of claim 28, which is *Methanobrevibacter ruminantium.*

30. A conjugate molecule comprising the polypeptide of any one of claims 1 to 10.

31. A conjugate molecule comprising the polypeptide of claim 3 or claim 8.

32. The conjugate molecule of claim 30 or claim 31, which further comprises an anti-methanogenesis compound, a signal sequence, an antibody or antibody fragment, a peptide nucleic acid, an antimicrobial peptide, or an antibiotic.

33. A fusion molecule comprising the polypeptide of any one of claims 1 to 10.

34. A fusion molecule comprising the polypeptide of claim 3 or claim 8.

35. The fusion molecule of claim 33 or claim 34, which further comprises an anti-methanogenesis compound, a signal sequence, an antibody or antibody fragment, a peptide nucleic acid, an antimicrobial peptide, or an antibiotic.

36. An antibody or antibody fragment which binds to the polypeptide of any one of claims 1 to 10.

37. The antibody or antibody fragment of claim 36 which is polyclonal.

38. The antibody or antibody fragment of claim 36 which is monoclonal.

39. An isolated φmru phage comprising at least one polypeptide of any one of claims 1 to 10.

40. An isolated φmru phage comprising at least one polypeptide of claim 3 or claim 8.

41. An isolated φmru phage comprising at least one polynucleotide of any one of claims 11 to 20.

42. An isolated φmru phage comprising at least one polynucleotide of claim 13 and claim 18.

43. A pharmaceutical composition comprising the polypeptide of any one of claims 1 to 10.

44. A pharmaceutical composition comprising the polynucleotide of any one of claims 11 to 20.

45. A pharmaceutical composition comprising the phage of any one of claims 39 to 42.

46. A method of inhibiting a methanogen cell comprising: a) optionally, producing or isolating the polypeptide of claim 3 or claim 8; and b) contacting the cell with the polypeptide.

47. The method of claim 46, wherein the cell is *Methanobrevibacter ruminantium*.

48. The method of claim 47, wherein the cell is *Methanobrevibacter ruminantium* strain M1^T (DSM1093).

49. A method of inhibiting a methanogen cell comprising: a) optionally, producing or isolating the conjugate molecule of claim 31 or claim 32; and b) contacting the cell with the conjugate molecule.

50. The method of claim 49, wherein the cell is *Methanobrevibacter ruminantium*.

51. The method of claim 50, wherein the cell is *Methanobrevibacter ruminantium* strain M1^T (DSM1093).

52. A method of inhibiting a methanogen cell comprising: a) optionally, producing or isolating the fusion molecule of claim 34 or claim 35; and b) contacting the cell with the fusion molecule.

53. The method of claim 52, wherein the cell is *Methanobrevibacter ruminantium*.

54. The method of claim 53, wherein the cell is *Methanobrevibacter ruminantium* strain M1^T (DSM1093).

55. A method of inhibiting a methanogen cell comprising: a) optionally, producing or isolating the phage of claim 40 or claim 42; and b) contacting the cell with the phage.

56. The method of claim 55, wherein the cell is *Methanobrevibacter ruminantium*.

57. The method of claim 56, wherein the cell is *Methanobrevibacter ruminantium* strain M1^T (DSM1093).

FIG. 1

A

(A)ATAATT AATTATT

B

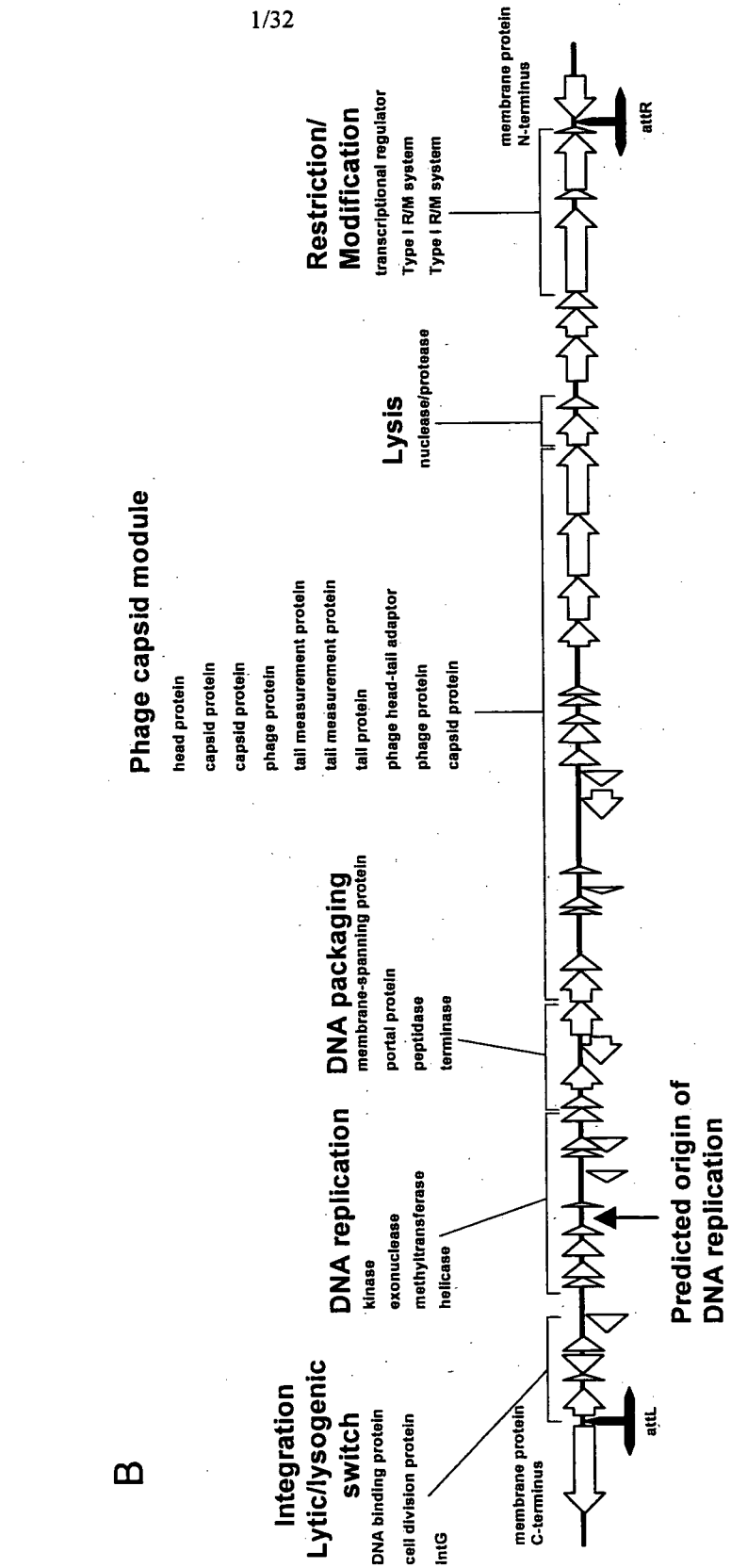


FIG. 1C

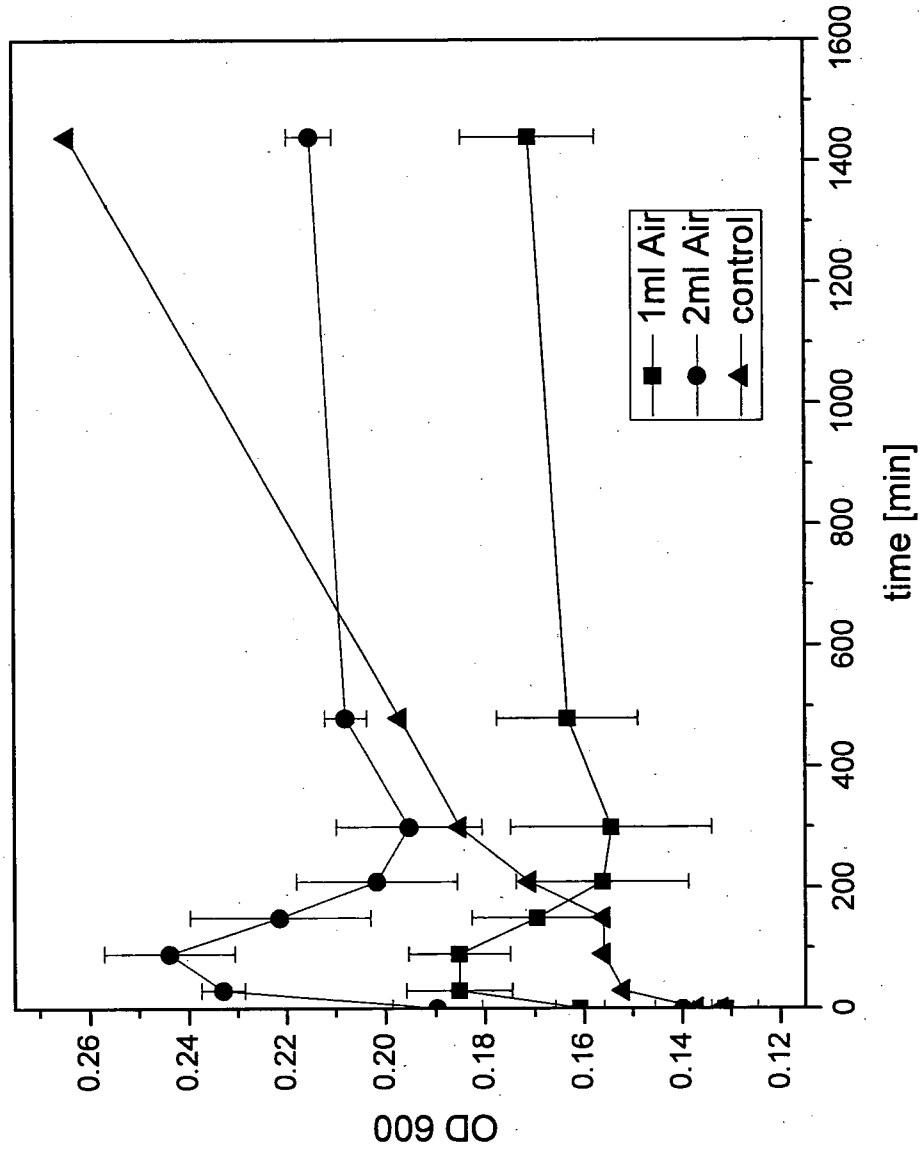
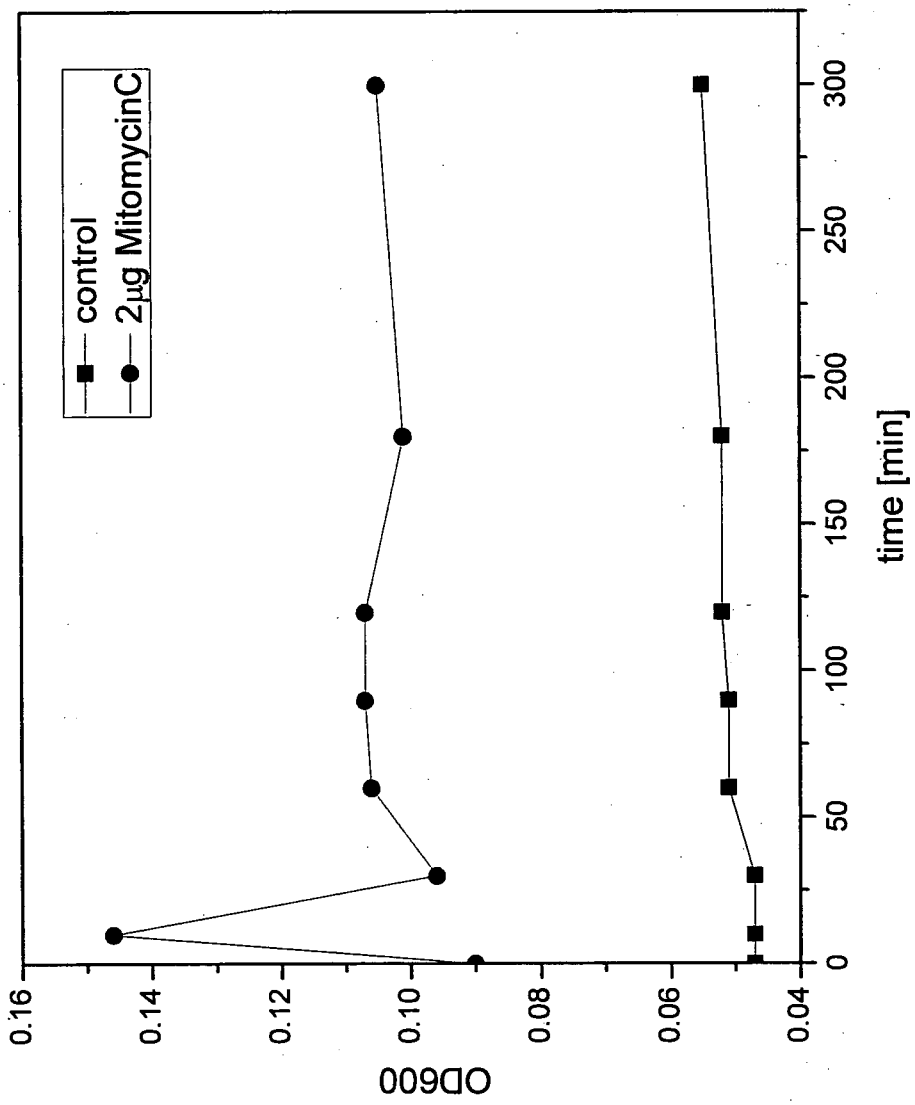

OD 600
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.26
time [min]
0
200
400
600
800
1000
1200
1400
1600
1ml Air
2ml Air
control

FIG. 1D


control
2μg MitomycinC
OD600
time [min]
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0
50
100
150
200
250
300

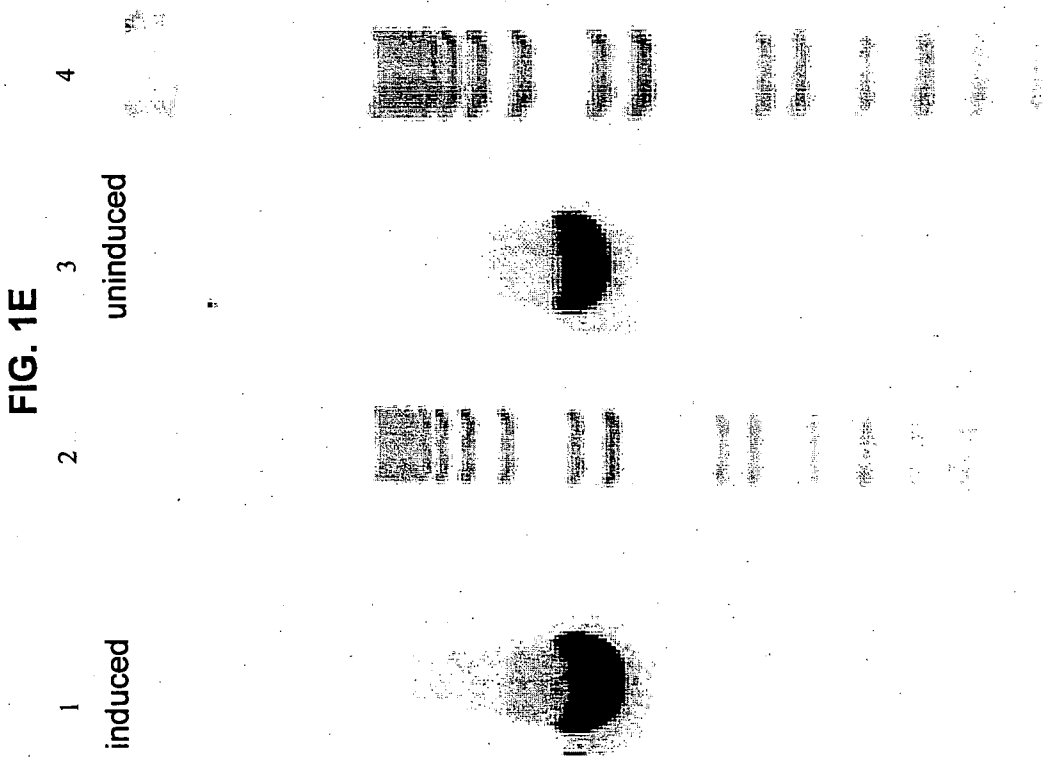


FIG. 1E

FIG. 2

Prophage φmru open reading frame annotation and comments

ORF number	Annotation	Comment
896	potential abortive infection mechanism, CAAX amino terminal protease family	phage defense mechanism
897+898	transcriptional regulator	likely to act on ORF896, likely frameshift
1451	predicted phage related site-specific integrase	1450 shows similarities to a DNA primase and 1452 reveals similarities to phage proteins and genome synteny suggests gene cluster, overall appearance of a mobile genetic element
1572	potential abortive infection mechanism, CAAX amino terminal protease family	phage defense mechanism
1578	potential abortive infection mechanism, CAAX amino terminal protease family	phage defense mechanism
1579	potential abortive infection mechanism, CAAX amino terminal protease family	phage defense mechanism
1783	phage structural protein	many genes in cluster 1776 to 1789 reveal similarities to phage proteins. 1989 is located at contig boundary, which could indicate a continuation of this phage remnant.
1774+1775	similar to GTPase subunit of restriction endonuclease	both genes resemble part of a (phage-specific) R/M system; the gene-cluster 1772 to 1775 predicted as involved in phage defence, based on gene synteny.
1959	potential abortive infection mechanism, CAAX amino terminal protease family	phage defence mechanism

FIG. 3-1

Prophage φmru open reading frame annotation, predicted function, and comments

GAMOLA ORF	Annotation	Predicted Function	Comment
Flanking genome region, genome integration locus			
1980	asn/thr-rich large prot. family	cell adhesion	by phage integration disrupted surface protein. N-terminus corresponds to ORF 2069
	attL	core in bold	Tttttaattgattgt**aataattaattatt**ctgggtctgac (SEQ ID NO:143)
Integration Lytic/Lysogenic Switch			
1981	integrase	phage integrase	
1983	cons. hypo		
1984	cons. hypo	DNA interaction	likely harbours a zinc binding domain and part of lytic-lysogenic switch
1985	cell division control protein	DNA interaction	likely part of lytic-lysogenic switch
DNA Replication			
1988	cons. hypo	DNA replication	DNA helicase or exonuclease, similarity
1989	cons. hypo	DNA replication	
1990	unknown	DNA replication	
1991	unknown	DNA replication	
1992	unknown	DNA replication	1 TMH (Transmembrane helix) predicted, likely membrane protein
1993	exonuclease	DNA replication	similarity to exonucleases
1995	cons. hypo	DNA replication	
1996	cons. hypo	DNA replication	
1998	ATP-dependent DNA helicase	DNA replication	ATPase domain involved in replication control
1999	cons. hypo	DNA replication	

FIG. 3-2

GAMOLA ORF	Annotation	Predicted Function	Comment
	terminator structure	resembles origin of replication	AAACCTTTATGGAAG TCGAGTAAGTATTGGTACGTATTATA AGTA TAAATACGTATCCTATACTTACTTTA TTAAAATTT (SEQ ID NO:144)
2000	cons. hypo	DNA replication	
2001	unknown	DNA replication	
2002	cons. hypo	DNA replication	
2003	unknown	DNA replication	
2004	unknown	DNA replication	
2005	phage associated protein	DNA replication	
2006	cons. hypo	ANTISENSE direction	adjacent terminator structure to 2007
2007	cons. hypo	DNA replication	2 TMHs predicted; likely to be a cytoplasmic membrane protein
2008	permease	DNA replication	2 TMHs predicted; similarity to permease superfamily; 2007 and 2008 potentially a frameshift
2010	ParB-like partition protein	DNA replication	cleaves single stranded DNA and nicks supercoiled plasmid DNA at AT-rich regions
2012	cons. hypo	DNA replication	
2013	integral membrane protein	DNA replication	5 TMHs predicted
DNA Packaging			
2014	cons. hypo	DNA packaging	terminase small subunit, in competition with 2015
2016	terminase, large subunit	DNA packaging	
2017	cons. hypo	ANTISENSE direction	
2018	portal protein	DNA packaging	
Phage Capsid Module			
2019	minor head protein	structural module	head morphogenesis protein
2021	cons. hypo	structural module	
2022	phage associated protein	structural module	also similarities to ATPase

FIG. 3-3

GAMOLA ORF	Annotation	Predicted Function	Comment
2023	unknown	structural module	
2024	cons. hypo	structural module	
2025	cons. hypo	structural module	1 TMH predicted
2026	unknown	structural module	
2028	cons. hypo	structural module	
2029	unknown	structural module	
2031	scaffold protein	structural module	similarity to scaffold protein domain
2032	cons. hypo	structural module	
2033	unknown	structural module	
2034	unknown	structural module	
2035	cons. hypo	structural module	
2036	unknown	structural module	
2037	unknown	structural module	
2038	phage head-tail adaptor	ANTISENSE direction	similarity to adaptor domain
2039	phage related protein	ANTISENSE direction	1 TMH predicted
2040	phage capsid protein	structural module	similarity to structural protein
2041	major phage capsid protein	structural module	
2042	cons. hypo	structural module	1 TMH predicted, potential signal sequence
2043	cons. hypo	structural module	
2044	phage related protein	structural module	
2045	cons. hypo	structural module	
2046	Cna_B collagen binding surface protein	structural module	predicted in phage docking
2047	cons. hypo	structural module	
2049	phage tail measurement protein	structural module	similarities, potential frameshift with ORF2051
2052	phage tail protein	structural module	phage tail component

FIG. 3-4

GAMOLA ORF	Annotation	Predicted Function	Comment
2053	minor phage structural protein	structural module	
Lysis			
2055	phage related surface protein	Lysis	similarities to Intimins/potential role as holin based on synteny
2058	peptidase		prediction as lysin based on location/synteny
Lysogenic conversion			
2061	dnd system-associated protein 3	DNA sulfur modification system, dnd (degradation during electrophoresis)	
2062	dnd system-associated protein 1	DNA sulfur modification system, dnd (degradation during electrophoresis)	
2063	dnd system-associated protein 2	DNA sulfur modification system, dnd (degradation during electrophoresis)	
2064	type II restriction m6 adenine DNA methyltransferase	methyltransferase	
Phage integration site, genome integration locus			
	attR		Cagaataaggat**aataattaattat**attgtttttattttttt (SEQ ID NO:145), part of terminator structure
2069	asn/thr-rich large prot. family	cell adhesion	by phage integration disrupted surface protein. C-terminus corresponds to ORF 1980

FIG. 4A-1

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	Flanking genome region, genome integration locus	
1980	gtgagaaatatgaagaataagagtttaatattaatttctttattattactgattacaataataagcataggatctgttgttgcaacgga taatgaagaaattaatatggataatataaataatattgataataatgaggatatcgctaatattgataatgtcgataatgtcgataact ccaatataaacaatccaactgacataagaatagacaattcaaacctaaatagagaaacagaactagattcaaatttaaataaatct	74
	Integration lytic/lysogenic switch	
1981	atgaagacttttagacaacagctattggaagaccctgaattccagaactacctgttgcagagaccgaacctgacggagagcagcctgca gtcatacctcaatgccgccaccaacttcgtgaggttcacaggggagccgttctacaagaccgtgcatgagctcagaagccagcagaacg ataggattgagaacaacatcatcataaggttcaacccgaaccagtcaaggataaacatcatgcagtttgagttcata	75
1983	atgaatacattaaaaatagaatgttcaaaagatgattatatcataaagactgcaaaggcaaattctgaaaatgagttgaaagttgaagt tcctgagaactggaactgcgactatgtgaatgctgtcctgtgggaagaggacatctgcgaagtgcttgaaaagggtgatgaaagaatgc ttttgattcctatgtgcggtgaactgcttcttgaaggagtgcaggaagatgaatacataaagtacatttgccttcct	76
1984	atgggaatgaaaaaagttaagatcgaaagtaaagcaaaaaataatatgacaagaaacgaaaaactcttctataaattatatgacgattt atataatgaagatagattaattttattctccacatattttaacatatatgataaatttatttttaaaaaagatataatccattatgttt taatgaactactctgaaaatgaaattattgaagcaatgaaaaaaattgatgaaattcagtcagaaggcattgatatt	77
1985	atgaaaacacaagatctaattaatataataaatgatgaggaatctcctgtattcctcaacagggaagtctttgagatggactatgtgcc ggacatctacaaatacagggacgagcagctggcgaaaatggcgatgtactgcaattcaatacctgacaacatagctcccaagaacctgc aattgtgcggaggcaatgcgacaggaaagaccacaacattaaagcagttcttcaagatgttgaacgaggcttttcca	78
	DNA replication	
1988	atgtttttggaaaaatgcgatggaaaagatacaatcagcatgtctcttcaagaaaagatgaatcttatattggaaactatggaaagcaa aggaagccccttcatatcctgtattcattgtaacatacctataaatgaagcagaagaatggtacaaaaatggagaaataggagatcagg actttataaatttctatgatgatgtaaacttgattgaagaaagttttgggtttgaaatttataaaaaatcagaatat	79
1989	ttgataattttccttaaactaaaatttggacgtgtaatcatggaaaaattaatcgaaattgacggagttcaatacacagaagcccaaat cagaagggctttagccattgaacgagacgtgcgctctcctaatttcgttgatatgttattgggaaagatcaaacctagcgaacttgcaa gcagagtttctgaaaaaggtgatgcctaa	80
1990	atgtgctatgtgggaaacaccaggacattggtctatcacacagaggactgcttctgcaaccactggctgttgaacgagaacaagaccat tctagaggaaaagcctgtagacatgaaaccttgcagcttctgcaaaccacagtttgacactgaatag	81
1991	gtgttggatatggttgcagagatggttgaaaacatcaggaaaggagaaggagatggatattccatctaccctccttttttcctgcattgt tttcctacaaggcaaaaagtattccgaatgctgctgcaaggctgaggcaagggatcagaagtttgcacttgtgaaccttgtaggcttca gaagagaggatgtcaaggtcatagacccaaggaccaacgaggagctgttcgtatga	82

FIG. 4A-2

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
1992	atgagcatgcttgcagacttcgagcctgcaaggctccacaagaggacatgggctgaaaggcatgatgtcgaaatccttgctgtcatatg ccttgccataagcattgcaatgctacttcttttctttgcgcttgcagagccgactgttgcaggagtgatttag	83
1993	atgactaaggaatttgaggacttcatgagaaggaacacaggactgcttgttttcttgagatgggacacagtggcatatctgaaatacct tgagtcccattatgacgaaaggaagtacgaatgcgcttacagattgttggaggcaatcgacaatctctttgacttctaccagataacct tcagcaaaaagactgaaagggaaattcctgaccagctttttgaaaaggacaagataaacaaaggcttcctttcacac	84
1995	atgaatgagatagttactacaactaacgagaataatgtacctgtcgacgttgactatgccatcgaggaatggaaggcataccagcgatt gaccagagagctattggacgaaacagattatcagacacacagaggcaggaagtacaagaccaagagcgcttggcagaagtacgcacgtg ccttcaacataaacacacagataatcgacaaggaaatcgtcaagaatgacaagggaatcgtcattgaggctgagtat	85
1996	atgattgagtgcatacaggaagagggagacttcacggattgggaggttccttcctcctcttccgacatcaagtacatagtttcagtgga tgatgagggaaacctgttctgcagctgtcctgacttctattacaggaagtccaggatgaaccctcacatctcaaaccctgaatcctatt gcaaacacatcagacaagtcttggaggaggacaatagattgcaaatgctctag	86
1998	atggataatataaataagacaaaaacaagtttagctaagtttgaagagtttttcagcactgtatacaaggacgaagtcatggaagtttt ggaaaagtatcctgaggaaaggacattggttgtggactatgagaatttggaaatgtttgatcctgatttggcagacttgttgattgaaa agcctgacgaagtaatcgcagcttcacagaaagcaatcaagaacattgacccattgatgaaagatccaaagttggac	87
1999	atggctaacaagattagagttaatctcacggttgaccctaatttgtggcaattagcaaaagacaagttaccttgcagcagaagcgaatt ctttgagaatcagttgaagatgttcttagggattgaagatgacgagtcggaaatcattaaggacatccaaaccaaagaaaatgagatta acgcactaagagacaagttatgtcatgttcgcaaatcaaagcagttgaaattggaatcaaacaaatctatggagaag	88
2000	atgacaattggaagtttggacaattttggaaagtcagatggagaaaacatgaaccctgaagatttcgactgttctgttttctttgaaat gtacaaggcacttttgagatattggatgttgaagttggcagcttcgctgaactcttggatgtgtacaagaatgtcgagatggattaca ccttgaagagacatgccctcaagcagaaggaaatcctttactggttcaataccgactggaaggaggaacttggaaag	89
2001	atgaatgtgaaaacagtcatgaatgaccttataggattgtcaaaggagttcgagggagtggaatatgagattgaatccaaaaattcaat ctacttctattcctttccaaagtacatgaaggaaggcattgtgattctgaaatattccgcaatctatgacctgcacaccatattgaaag gcatggatggaatcatagtggacattctggaggttgaggacaatcctggtgatgagaagaaggatttgctatatgtg	90
2002	atgattagcgacgaatgggaagaggaatactatgtcaaggtcaacgaggaactggagcaggtggaagtcaggtttttcagcaaggtcga caggctcgtctttgctcaatcctactcttcttcatcttccttttcctttgaggaggctgaaatcctgtgcgacaggatatgcgacatac tgacaaacgatttgggaaatgccaagtattatctaggaggaaaagatgaattatag	91
2003	atgaattatagggaatgggttgcaagccagttcgagctgaggatgactgaggacggaagggtttgcttcaagacaccatgcagctacta cacattttcaaaggaggactttgagataataagggagatgttccttaatttcgagtcctga	92
2004	atggatagaataaaggaattgaatgtatgcggaacctgcaagcacagccatttgattccagacatcaacggagagatagcagtgaacat ttgcaggataggttcagaggctgtcaacaaggacggaggacttacatattgcgtagactggacaccaaggaggaagctgtcatga	93
2005	atgttaagtaaaaaagaagccatacagatgacgctggacaatgagaagcactatcctgtt aagtgcaagtactgcggaaagccgttcaccaagtctcacaacaggcagatgtactgttca gacagttgcagacggaatgccttgagggaacagaaggcaagataccaggctaaaaggagg	94

FIG. 4A-3

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	ctaaaaataaagcagaaagtgctgattgtagatgaatacaagaaatacggtttgggtagc tatggaacaagtgcgaacggacacagaaagaacaatttttcacatgagtacatggctatt caaaaggagatgaaaagaataggattgaagagataa	
2006	atgtatttggctaaattttgtcctaactgtggaaataaagtagaagaaaatgataaattctgcatttattgtggaaataaactaagagt tataattcctgaaaaaaaagtgaaaagaagctcaaatagtattaatgatgaaaaaactagcaaatatgttgaagtcattgatgggctaa tgagatataaagtttttccatcaagtttacctgtgaaatatattatttataaagtgaattatggaacaacatctgatgaaata	95
2007	atgatttatgataaagcgacagttacaagcctcatagtggctatcctattgccattgatgagcatgttaggaatcggagaattgactca gaactacatattggcaatagtgagcggtatgatagcccttgtggtttggtactacaacgagaagcacaacagcgatttggtgagcggaa ccaccaagtgcgactgcgagctttgctatgggggagacgatgaagcacttatatga	96
2008	atgaagcacttatatgagataatcccttacaggaggactgtttggattacagggttcctgaagacaacagtctcctcggcaatgataac cactggcgtagtgatattgttcaacagcataaccgagcacccttacttcatggagtgggacgagataggaatcgttcttggaatcgttt ccatcaccatcgcctgcatttacatagcgatgatagacagatggaaggaacgcaggaagaaggaggagcttgacacg	97
2010	atgatagaaatcagcaccataaagatcaccgacataaagcctgccgaatacaatccgaggataatgagccagcttgaacacacaaagct caggaactccatggagacattcggagtggttgacccaataataatcaacctgaagaacaaccacatcataggagggcaccaaaggtacg aggtccttctggacaagtccatggaggacaacgagttcataaaggagctccacctcatcaggcttggagatgtaggc	98
2012	atggacgtcgacaacctgctgggatatgacaaggtactggaattgaggtctttattggaggaggtcacggacaaggtgatacccgtatg gcacaagaaccgcggaatcaaggacttcaagcagatgtgccaggactacaatttcgtcagcataagcggctggagaaacgaggacgtca aggatgaccagttcatccatttcgtcaggcatgcacacaggaacggctgcaggattcacggattgggattgacacggagaaaggttctc gacagggttcccttcgacagcgtcgacagcagcagctggctccagaccatattgtatgcaagg	99
2013	atgattaatgaaaggataagaccttatttgaattttagcttcaatgacaagaaagtcattttcgtcacattgttcgttgtgagcaatct gataagcaacctgttagccatcaaggttttcaacttgggattttggggattgacaaccgattgcggaaatctcctgttcccgttaggat accttatggcagacgtgattacagaagtctatggagagaggacagctcgaagggtcatattgcttgggctctttgca	100
	DNA packaging	
2014	atgcccgaaccttgggaaaggcaaagggacgagaacggaaagcttgagccaataaaggcattcgagtacttcaccgagtacctgacaat ggacaagccacgaagcatgagggttctgtgcgagaggctcggcaagaaagatgggtatattaggcaacttcatgcctactcatccacat ggaactgggtcgaaagggcagaggcatacgatgaacacataatattgaagaaacggttaaggaaagagaagttttac	101
2016	gtgaaggacattgtcaatcattacggctacctttcaccgtacaagccgaccatcaggtctgatagtaaggctaaaaacaagttcaagct caatgagccttatcgtggacagatgttgagtgcaggtgctggcggttcaatcatgggatacggtgcaggtctccttattgtcgatgacc ctatcaagaacgtcgctgaagccgaatccaaggtacgtcaggcaaagctcaaggactggtggggaggaaccatcaag	102
2017	atgagtaagaaacaagaaatgatgagaattgaaaggttaaaacactatgcatatcaaacaggactaataatccctatttttaaaaatca tgatttaattaaaaaaatagaaaatggaaaaataactaatactgacgaaataaagatatatattgaggaaaatgaaaagcagataaaaa gaagaagagaatttataagtattatttatgataattgcaaatattttaaatacgacagcatatgttataaattaatatccaaagtaaat	103

FIG. 4A-4

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	aatttt	
2018	ttgagtaagaatgctaaagcagacgcctttgttgtgaccactgaagacggatcctatgacattgtcgatgcagatgtcttggagaggta tgcaataaagagcgaatctgacgagacaggaagcaagcagttgaagactgatggttgggaatatgatgatacattgcttgagccgttgt atgacccattgcagttgtgtgagctattggagataaacacatatcatgagaactgtgtcgatgttgttgcaagggac	104
	Phage capsid module	
2019	atgatggcgctcaaggtgcgtcattcaaagaggcagattcagaatatcaagagagaatacagaaggcgtttgatactcgaagagcagtg cagtagggagattgcggacttcttcagaaggcttgagaggaaaatccacaaggtcatggatgagcactgggaaagcgaattgggtcttt tccatctgaacaaggtttcagacatcatacaggacagcaggcaggagtactatgacatcctgttcaagtactgc	105
2021	atgataattgaaattcctggaaatgacaggacagaagaaataaatttgccgaacggacagtttgtactgataacctacttgcaggaaaa tgacatgatgtcattgcctgacgggaagtatgtctgcccattcagaatgattcagcttttcactgaagagggcggtgaactgattggag agtgcattgaggaaaatcctcactacaacactcgcttctacaatactgtctttgagcatttggatgaaggaattgag	106
2022	atgttaagagtcgtagagagaacctattatcaacaagaagaaatcaagacactagactgtcgaattagagaggcaggagtcaacactta ctcccttgctcgtcaaggagcagttgactatcctacctacgaatcctacaacgaggtgagagaggaaaggataaaagaggctaaggaaa aatacggtgaatcctactactatcactggagagacgttgagacctacttctattacttaggaagattctttagtgat	107
2023	atggacgcgataaacgttatcaaccaaaacaagatattagtagatgttttgtaccgtgggactgtaaacctgattgacatcgtcatcgg tgatgcattggtctatgacaatccgacagttgtagtcaagtgctacaccacagatttggcttttgcaaccacagagattgatgagattg tattagagaacgaagaatttgagttattaataacttacgaggacggagagttctcaattttgttgaagtcccacaat	108
2024	atggtttttgtcgatttgtgcgagaacgaaatagcagacatggtcgagagcttctatcgcaatggagacggtagcagcaggatattgac caacagagcagttggagagatagtggagcactactgttcctttgaagtagacggtagaatcaccactaatctccgagatttcctattat atagtgttgtcttatatgacaccattggagaggctgtcgatgacggagtcaaccttgaggaagtgttcatcatagaa	109
2025	gtggtgaattggttgaaggctattggagataacttctcagtggattatctccttttagcattattctctagtggagatttgatattggt tgccatcgttctcaattcttatggtgtcatctcacctgagaatgtaagggaactggtgattgactatataagttacaggaaagttgata ttttttggagacatctaaggagacctagaatgagttttgaagattatgtcttagacaattttgaagagatggaaact	110
2026	atgacaatgaaagttacttttgaagatgagaacggtgagagaacagtcgagttcggcgatgacgtcgatttcgttttaatcgaaagcga tgatgacggaaacattgagattcgcgaaggagactgggaattggacggagattccgacgatgattgggaagagtacgatgactgggacg aggaggagttctga	111
	gtggacgaattcgttgagaccctgttcgacacctactggaaggtgaatgagaacggagagtacatgagccttaccgattgcggtgattt ctacatcgcaaaggttgctccctgtgttcgcaactggagcatagtgatagaatgcaattgcttctgttttcattgcaaggaatttgtct accatgagaacggagcaatacttgagataggtatggagattagctctctttacctaagccaaatggagataaaagatttgaagatatac atgatagattgt	112
2028	atgtttttggaaggaatatttgaacaagacggagagaatgtaagagaacaggtgatttactggagaaagtccaatcaggtacacaactg gttcgtagtcaacgctcaggacggagaggacaactgccaaccacactcagtaagcagagaacagttagaggaactaagagacctatgca	113

FIG. 4A-5

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	gagcagtgctcgcagacaatgacaaggcagaggaacttctcccaacaagacctggtttcttctttggtgcgatagac	
2029	ttgcagttggtggtagaaggtgagaatatggaatgcccttgcgataattgtgagatgttaagagagaacgaaccaatcaaggtgattgactggaagtccaattctccatttggaaatgggatttktattcacttttggagggaacggcttcaggatgttggctataatgggtga	114
2031	atggaactgatgactagggaaatggaagggaaactcaaatccttccctttctattcccaagacggaaaaggagatgacgcgatagtggtgatgaagttcttcaacccttacggattgggaacttggtatgtattggaggcagagaaacaggagaatggtgattacctttttcttcggatatgttgaatcaccgataacacctgaattcaacgagtacggttacttctcattgtctgagttggagaaccttaagata	115
2032	atgttcggacagaagaaagaatttgtcaaaaagatgtatcatgtcggagacgttgtcgaactggttcatatggacgatgcacaggctccacctagtggaaccagaggagaaatactcttcgtagacgatattggtcaaatccatgtgagatgggagaacggctcaggtctggcactaatatatggtgaggacaggttcaaagttgtagagaggaaaggagaatag	116
2033	atggatttgttgtatctttatgatgacctgactgcaaggagagaagtctacgattcagtcggattgagttttgtggtcaagtataaattctcatctcgtaaggaagcacaggatttcgcattgaagtacggtgcagaactgattgaggaatga	117
2034	atgattgaacggaggagattaggaatgaagtacgatattttactatcttggatgaaatttcaaggaagttagatgacggtgaattatccgatgagcaggttgatttcctgttgcagatggagattctagtagaggaaggcactatcacagacgagcaggcacaagatgtcatgaatggagattattaa	118
2035	atgatgaaattaagtttgaaggaactcggtgaggaaattgagatgatacttgcagaaggtggattaacctttgatcaagtcgattatctgttgtacttggagacctgcattgcagacggaagcataacagaggagcagaagagagaaattatctgtagggacttctaa	119
2036	ttggtgagagaacaggagagattgattatgtatttggaaattgacgaggtcaagtgcgaaaacattgacagaattgagtttgacgatgtcgcaatggaaatcgtcctgacagacgaaaaggtgtacgagagaataaagagatggctcaagtccaatgagattgactacgattgcagggaagaccggtatttcgccaatctcatagaatatgtcattaggataacttggtggtaa	120
2037	gtggtaaccgtgagagatgtcggattcaccattgaggaacggttcttccttactgctcaggaattggagtattccgaagtgggagaggaacacgagagcgtcattgacagagccatcgcattgctgtacacgaagctcagaaccagagacttcgagttcaccgacgaagaacgggaactcttggaggatgcttttgtcattgtcagcgaccaatag	121
2038	atggctaaagaaaatgttatagattataagattgagcgacaaaatgataacacttggagttatttatatgtaactgaaagagggagaggaaatatcattgcttcatcttttggagaattgaggcagaaagttttaaaaagaggtttgccttggaatgatatatctaacctattcactaaaagagttctgattctaatgttcgtcgcgaatctgtaaagaatattgatgatgaatctgttttggctgatgtcgct	122
2039	atggtagtcaaatgtcctaattgttttagtccacgtgtttcaaaatgtgaagatactaatattaaatggcaatgtgataaatgcaaatgcaaatttaaccatggtgctttcgatattaatgctgaaatggagaaagtagaacaattaacaatagaaaaaattgaaagagaacgagaaagaactgaacaatttgaaagagctattaaagaagcaaaggaacaatttgaaagagaaagaactgaacaatttgaaaga	123
2040	ttgaccgagggtcagataaggcagattgcacatgagtatctggcaaattacagccttgtcgacaagaaccatgagttcttcgagaccagagaagttataggagttcctgtggagtcatatataacaaacgagcctataagcctcaagggcttggacggaacagtcaatgagtatccaaaaggaacctggatagccaccacaaggataactgatgaagaggagatggaaaaggcactcaatggagagtacactgga	124
2041	atgggaaacgaagctactttaaaccaattggtaaacgagcaggaaaaggcagtcttcaagtccatgaggactgacatggagacaggaaaggctgtattgaacgtcgagcagctaggttacttccttagagaagcaacattagacaatacaattttaagagatgcggacttcaagctga	125

FIG. 4A-6

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	tgaaatcattcaagaagcatctcaacagggtaggaataaacggaagggtcctcacaaacggatatgacgtgaacggcgagaccgaccct gagattcctgcggctgacgtcgacttcggagcaaacgagttggacgtcaagaagctcaaggca	
2042	atgcggaggcgttgcctgaactctccagagcataatggtatgatttcgcatatcattgtgcttctcatttgctttataggattggttga ggcgatactgatggcattggttgattgggaggacttggcaatatccgttcgcaagtctcctagaaagctttataatgttttgaaggatg agttaggtcttcctgaatggaacgaactgtctgtgattgaaaggagaagcatgaagaaaaggtatgctgtcataaga	126
2043	atgacttggattggtacggaagatgtcatcgagttcacaggagtcaagcctcagacattcaggttcgagaagggagacacctccagcct ggaaacattgcttgagaagtggattctgcaggcagaaggacttataatatcctactgcaactacgatttcaatgacttggaggagatac ctccagcagttgttaatgtctgccttaggctcactgcaaacatggtcgcattggcacaggcaaggaaggacactcct	127
2044	ttggtgaaattgcagattgacgttgaggaactcaagccattggagccaaggttcaaaaaggttgccaaaaggacagttgtgttgactgc aaatgaattgcagagaaacctcaagaagttgagtcctgtggatcatggaaggcttcagggctcttgggtaatcttccagacaggagaat tggaaaggactgtgaaaagcagtgcaaagtatgccattttcgtgaatgacggaacaggactgtacggtcctttgggtcat	128
2045	atgagattcgtaaatactgcttcgcttgttcctcagactgtcaaggcatatcttgaaagggaaatctgcgaaggagggttgcttgagga tgtagagacactcattccgtccgtgaacagcgacgttcctgttgacccacctgcgatatggatagtccagcaccccactaccagatggt ctggcagtcagccaaatctctcaaacaagatagctatgtcagtccctttcgagttcgtatgcgtggaatacagcgatgacttggaagag gctgagatattgggaataagcctagccagcagggttggctcaagcctgatgaagaacttcaacaag	129
2046	atgggaattcgtgttgtaggaatgaaggaagaggcaaggtatggagttgcggagtcagcgccggacttccatcaggaagttagcaaggc aaaggcttccttgaactccactccgaacacaaagtcaagcggctcaaggatgaagaagaaggcacgtgcaggcgtgtacaagcctactg ccaacatcgaaggtgaagttgacttgaagaggataggacattatctcaaggctttcctggacaattaccattttact	130
2047	atggttgtagttaagaaatccgatattttaaagggcgtaaaaaagattgaaaaggtgaagattgaggctttggacggagacgagatgta cttgagaccgttgtcccaagccgagatcaacgaggtcgacgagattgaggcaaaggctatgggaatcttcgagaccaacgagaccgcac acaggggaagaaggcagaagcctaagagtgtggttgagagcaaaggaaagataaatctcgaattgcagcagaaggca	131
2049	ttgccgtcaagcaatgtaatgaacataatagtcaaggcagaggacatggcatcatctgtcgcccaaaaggttgaaaacagcttcaggaa attgggaaacacaatagacagcaccttcaccacatctctttcaaacaccaagttcaatcaggaactcacttcttttggaacagacttgg acaaggtaacccaaaggctgaagcaggtaggcgtgaacggtcagtccagcttcaaccagcttacaaatgctgaaagg	132
2052	atggcaagtgttacaaaatatccgagcaacgtttcacagaccactggaggaaagttcgtgtccttcagcaatctggcaaacataaagaa caatgctgacggagcgcatgccgtgagcagtgttcttatcaaaagcaagaagcagtctccaaacagaccgtccacagtctcatgcaaag gcttcggattcagccttcctgaaggtgctgaacctactaaaatcacagtaacttataggcataggaagaatgctggaagcgactacagc tcaaagaacaagactcacatatgcaacatcggaggacctacaatc	133
2053	atgggcatagctattgttgttatggacaatgaggagaacttcctgcaattccttgaccctgatttatgcaccatcaatgagaccataga ggaattgggcttgaggactttggagttcaactacaagtttcaggactatgttgaggacagggatcttttcaggataggtaacaagatct ggattagcaattcccagagcttggaggactgcctttatgtgataaacactcctgtggagaacagcgtctatcaggag	134
	Lysis	

FIG. 4A-7

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
2055	atggtcgagaagattacagtgagtcctcaggaagtcagaggatacggaaatgttgttgatgagaaggaattggaagattacggaagcta caggtgtgatgtgagcgagagctcggaggtgatcaagggagttgaggaaaggatattcagcgtttctggtgttcctgctcctctgga gcattgccaatgttactgaggacactcgcagggggcaggtgcgcccacatctccgcctcatttgaggatgggaaggc	135
2058	atggttagattcagcagagacatgctccaggacggagcgaagagaatgttcaagtggctaagaaagggcgaagggttgcctaactactt gataatgtatgacatggacaggaataaggagtataagttggttccaaaggaatatgcaggactgtatgagtccagaaacatattctgga ttaagaacggaagggagcctaactatgttacactgacttccgttgcaaggaatcctcttgtgatggactaccagaacaccaattacacc tgttgcccaaccagtttgtcccttgcctcacaaatgctatatcactatgtaagtcagaaagtgaatgcgctaaggctttgggaactagcaa gggcagtggaacaagccctgcacag	136
	Lysogenic conversion	
2061	atgaaaaaacattgctttttatttttaggggacagtttgccgatatatgcaatgaagcc atgttttgtgaaaaacatttagtagaaggaaactatttggattcaattatccgtgcagga aaggcttcagagataataactgtaaacattgtgaacttgaaggtcaagatggcttaata agctctggtcagaagaaaaagattggaaatgcttggatataagggtatcatttcttatgat atttataaaagattaaaccatattcgtaaaattagaaataaggctgtccatgggcatttaagtgcattgatattgaagacaatgcaaatattct gcatgcctatctgtatctaatttgcgcatatatttctatagaggaatatagagacactaatttttcagcggaagattatacgggccctatta tggacattgcctctaagcctaaagagactgcttcagagacttcagaggataatgaaaatattggagagtttatttcaagtccattggat gattatctttttgaaaagtatgatgacagttacctgttaaatgaactgtctaaacttaaggactcttcaaaggaagctgttgaagacga taacttaagcgaatttaaggaatatctttcatattgacagatctattcaagaagatttttttaaaggcattgaacagagccactagtttta attcatctcatctgattatgctttgcggtagtgttggagatgggaaatcacacttgattgcaaatttaaaaaagaaaaaccctgaactc tttaatcaatttgctatccattatgatgctacagagagtttttgatcctgaaaagaatgcaatagacactttagcctcagtattggaacc ttcaatgacaataatcttaacaattcgacagaaaaaacttatttttagccattaacttgggtgtattgaacaatttcctggaatcatctt atgctaatgaagattacactaaactcaagttaattatagaagaagcgaacatatttgaatctaatgaggtttcagataacatttatgga gataaggttagtttttgtcacttttcagtgattacaacatgtttgaattgaatgatgaaaattcaaactacacatcatcaaagtacat ctctttccctatttaataagataacccaaaaggaggatacaaaccctttttatgttgcatatctcaaggataaggactctcactttatca atcctataatctataatttatgagatgctgatggatgaggaggtccagaagacaatcattgatattatttaatcaagatttttcataaaatac agaaaaatcatttcaactagagacttattgaacttcatttatgagataatagttcctccggaattcctgaaaagtgaggatttggacaa tatcaatgatttcatggactatccattgcctaatttactcttcggatatccagaaaggtcagatttattgaagctatgcaatgaattgg atccgacattgcatcgtaatgaatctttagataagttcataatagacttgaacattaatgacgatactgaaaagatattaaatcgttat ttgatttcactcgtttaatttccttgaagagtatctggttgatttttagggagtttaacaattctgaaaaagagaaggt caccaatattctaattaggtttgcagtattctatggaaagagcattatcaagaataatttcaaggataaggtttattttaaattatttaa aatatctgtatgcctataatacacagtctcataaggattacaaatatctgttcactgaagttaaggacgctatttttaattggaaaggt tcctataagaagaatactatatgtattgacactttggattcatttaaagtgtataaaatttaaaatttaaaaccatctgttgataaatt	137

FIG. 4A-8

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	tgaaaaagtttattagatggtctcttttaggaaatagatttaaaacagatataaaaatttatttttcagttgaatctaacaaaaaga aaataccttaaatgttgatttttcattatatcaatacataatgaaattatatataatggttttaagcctaatcaatcagataaagacgat ttaataatttttagatgaatttattaataatttatagatgaagatacagatgatgatttatatgtaattagtttagaaacttatgaaga attttatttgagtcaatgatttttggcacttttgaatttaagaggggttaa	
2062	atggatttttcagaaaattatataatattattaaaacaaatgacttgtgatgtaaataaa aggaaattaatacatcagattaatcaaaattctccattgttacctttttaaaactaatact cctaaaaaagctaatttttgaaaatggttttgatataaatattgggtgaattatccagaatc ttattaaataaaactattgaaaaaaatttttaaattagataatattgtttcaaatttaatt gataataatattgaaattgaagatggaactaaggagtatataacaaaattgctaaatgaa tatttattcgatgaaaaaaatgatttaaaaatatctccgaatttgtatttatatatt cctctttcaaataataaaagttctaatggagaacaggaagtagcattgttttaagagat attttgtaagaataatcaaaatctttattaactttttttgaatcatatgattcaaatcac attatcttaaatttaatttttaaaaaatactcctaatttacatcataaaataactgaaact aagtatgtaattcatttgaggaaattgcaaatttatttaatgaagatatcaattatgca atactttataagaaatttttttatggaaaacattggtaatatttttgcttattactatttc ttttatatctctcaattaattctaaaaatttctaaaggctttaatgataataatgaatttgaaaaaattgtactatttattggactggga atctgcaagtaaaaatcgtaaatcattaaatagtttactaaaacatcattctaaaccattatatgcaaaaatggcagtcatag atcaaataaatacactattaggtacaaataaatttattggaaaaagatatttcagaatatttttaatatttggacatcaattctaaaaat aattttttacatttttcttaaaaaaatgggtttcagattatatggtaagaatttgatgataaagaattaccagataatttattaga attaactgaaattcttttttgaaaagtttaaaaaatgaaaaactgggtggatggagctgtacaatctaggtacgctttaaatcttgaag atatagctaaaaaaatatttgttaaagcgtagaggttcgtatggttatgtattaaatattaatagagatatgttgcttgttttaactgca ttatgtgttaaagataagaaaattaagttaaatcaactatttatttgaatttgagaaaaagaggagtttatttttgacaaatattcaaaaga agaagtagtaaatttttttaactaaattttaattgacaagaaaagtgatagtggagatgctccaatatgttaaaccagttttatgat	138
2063	atgttaaaccagtttttatgattatctctaataaattactaaatttttgatgatactaagattttaagtggtgaaaaatttttcat tagttttgatgaagatgaccaaataatgtctttttataatagtttaagaagtattgcagaaactaatttttcttgttctgaattcatat atgtccatactatttcaggaaaggaatataatacttatttccataaacattaatggagtcaaatttgtcatttctgaaagtttaacaatt aacgtagatttcttagtttacttttctcaagagggtgtttggaaggatactgcattattagtttatctgtaatga agcaatagatagtattggaaaaggtatgagaaatttacagaaagaagaatgccattaaatgtaaaatctatttcaaaaaatttagagg atgaaattaacgattctcaaattcttaactattccgataagcaaatcgctaagtttttcttgaatattcaagaagaagaattgtttcaa acaacattatgggattatgaaactatcttatcaataatcaacaaaggtttttgttagtgatgaagatttaagagagttgaatctttttaa agatgaccaactaaatcaaaattctcctcaaaagatgttaaaaagattaaaagaaaattatgacacatttaatgaagtaaataaatttt cacaatatggagataaaaagaacagcttaaaaatatgtttactgatttctggagtttcttatttttaagtaaagatgactggtataaagca gaatggaaaaatggttaaaaaatcaaaggatgacttttattaatcaacaaaatcctttaaattacaatgaaaaatcttgaaaaaatcacaga	139

FIG. 4A-9

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	aaatggcttaaattattattgggaaattcctaactcatttactaaaactggaaaaagaaaacgaaatatcattgttttttaatccaaatcatt	
	ctcatgaagttagtttaaagttcagttttgatcagattcttagtaattcattcctttacatacaaattcaaaaaaattcactattgcaaga	
	ggcaaatcattaatagtcaatttttactctggattctagtgaaccaatttttcaaaactatataaatataaacatataaaaatgaaaatatctc	
	tgaatttacttttaatattgttgtttttgaatttttaattctattaaatctcgttttagtgtcaatgtcaaatcta	
	aacagattattgtaacaaatgatgaggatagtttgtttttggttactggttcaaaagaaatagagaaattaattaaagaaaat	
	ggtgaaaagtttatatctttatgatgaatctttaataatttctgaacaatctcctgcttggaatgatggaaaattaagttttaaatt	
	atataaagacaataaattatgccccatttttgattaaagaaaaacactgccagtaaattcctatgtaatttggaatttga	
	aaagaagaaatatggaaaacttttatttttaatggagtaaaagctgttcaagatgtaaacagtttctatctcgtagaagaatttaaagaa	
	tttctaaaaaatggaacgagaaatcattaaacaggacatatttttatgctaaaaggaatattgatggttctttagaaaaaattgaagtttc	
	attcagtaacgagttagaaactgcttacatggacatatttaattattataaagacttttgatgattcctccagaagacaatctttccaagtt	
	tagtttatttaaatgatgatttataaagaatttataactatcttcaataaggaaatatcagaaattgaagagaattct	
	atttgtctgatttttaaatataaaaagaatttgcttaagttaggtcgtattgaaactgataacaaaataatgtattctccattatctcc	
	actaaatatcgcttatcaattagaagtttctaaacaatgcggaaatgaggatttatccgtaaatatattagaacggttagttccaaata	
	acttaattccttatatttgttcagatgatggaaaagaattgtttagacctatttatcaagaagaagctcatgaatggttgatttatgag	
	aaaagtgaagaagtatcaattggaacaacaaatgtatttatttctaatgttgttactgaaaaacttaatcaatttgtcaaacatttaa	
	ttacttatttagttttaacaactcctcaccaattaaaattaacttaatcaatataaaagatgataaagaagttgtgaaaggagtattta	
	attttattagatctagattgcctgataaaactaaaacaaaaaaagttattccttgttgaaatcaatatttacaacgacgctgaaaagagt	
	tctttgacaatttatttgattgtcagtctgaaattcagttactggaagaatttggaatttaaaaaattaaaatcagatatttttgatcc	
	tatagatattatccatatgattcaaaataatattttcatattaaaacatccatttaaaaaagaagaatatgaatatgctccatcttcttt	
	tctataaagtcaaatcacatacaatattgcaaatgacaatatggataaaattgaaactggattatctttaaacggtttattatcttca	
	gttacatcaactactaaacattcagaatatagaacgggtttcggtacaaacaatatttttaaatatgagtaatcctcttattaaaacagt	
	gattaacttgaacgaacttgttgagaatagcaagaacttttggaaaaaatacttattcaaaaaataaatcagttatcactactgttgaat	
	tagaagaagacaatattgaagaattgtacgacaaatctcattgggtaacatttattgaacctacatttggtatcgaatactttgacagt	
	tctgatagtaatttgatcattattcattacagcgatcaatatagttcctctagcaagtatgatactattactgtaactaataaatcaac	
	acaatatgaggaaattataagagatttcctccaatctaaatatgtaaaagttacagatgaagaattgtatgatgtaataaagatgttta	
	attcaattaatggtgaatggttacttagagtaatttctaattccggccattatagtagggaaaaattaagtattatttctgctattaaa	
	tattgcttatccatttttggaccataaagatattgtttggattccagtttctatggaagaaatcttaagaattgcaggaaatgttaaatt	
	agataaaaataaaggaatttttgacttctaagttaataaaaggaaatcatagtgatgatcctgctattcattggtgtaaaattcaatgaag	
	ataatagaatcgaagttatattctacccaatcgaagttaaataggtttgaataatgcttcaactattaaaaaaggtaaaagtcaattta	
	gataatacttataaaactcttaaaactcaactacaaaatattaatgtagagaattctgaatttagaaataaatttctttagaaatttctt	
	catccaaatattattgtccaatgagcaaaaattagtaactaatcatatttgggatgaaaaaggattggatagaattgaagaattcaaag	
	cagaactattaaatgatgaatatgatatttttatatggtttagaagagtacattggtaaaggttcttttggtttcatttaaaaaagaatct	
	catcacatcctctatctatatggatgttgataaagcaagtaattgagttaccagaagatttttgcttattacggttttagccactcctataag	
	agaaatccatgatgagattcagtcagacaatacagatatttcttgcagaaacattgttatcacatgtagatataagtgaaattagagcaa	

FIG. 4A-10

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	agaataatgatatatgcgattcaaatgaggatatgtcaatagatgatgatttgataatttaagtgaatttgaggatagtttattgaa gaagaatcagaaatttctgaagagcctgatgaagaattaactgatgcaacttcaagcagtgataatgaatcaattgagaatattggaga atctccttctaaaatttctaatgtcagagcattaatttggaactcagaagggttataatcataaagtttattgggaatttggacatcctt ctttagcaaataggcatatgttaattcaaggaaaatcctggacaaggtaaaacatatttcattcaacgaatgttaaaagagttgtctatt caaggcattcctagcataattattgattatactgatggattcaagccttctcaattagaacctaatttaaagactctttaggagataa aataagccaatatttttgttgtaaaagaaaatttcccaataaatccatttaaacgaaacactatcatgattgataaggacatatttattg aagaagacaacagcactattgcgagccgatttaaatctatcattaattctgtttatgggctaggtattccaacagtccaatacttttat caaacagtttagattgcttggataagtatgatgataatttttgatttgaatatcctaaaagaagaaattctaaaagacgaatcaaatag tgcacaaacagtcttgaataaattaaatgagttattggataaaaatccgttcgcatccactgatttcgactggtcctgtcttagataata aagacggtaaagtttacattatccaattaactgcactttctaaggatatacagacaattataactgaattcatcctttgggatttatgg aattacaaattaacaaatggtagtgaagataatcctttcattgttgtttagatgaagctcataatcttgatttttccaatgactctcc atgtagtaaaattcttaaagaaggcaggaaatttggatggtctgggtggtttgcaacacaatctgttaaaggatctatgaaaattgatg aaatagctaaattagagaatgcagatgagaagatatatttccatccaacggatgtttcaacaatagctaaagatttgtcaaaagacaat gaagataaaaaaatatatgaaaaagagttatctcaattgacaaaaggatattgtattgttcaagggtctgcaatagattcaagtggcaa ttgtatcagccaaatcctgtaactgtaaaaattgaggagatataagtttgatgaaaattaa	140
2064	atggaaagtaaagacagaatagagaatactatttcaatgataaaagaatggatgcttat atagacaaatatttttgaagagtttactctgagttctgaacaagaagaggcactgaatatt tggatagataaattaaataacgatcaattaacaagtgaaaaagggaattatcataatttt tttgaaattatttcttgaagatttactttggttataaacgttctgatgttaaacatgaagag aatattggtgatgaaggccatcctgtagagtttgtattagaaaaaagatggaaaagattat gtgattatagaactaaaaggaacaacttataaagacctcactaaaagaagacctggacag caatcaccagtagagcaagctacaaatttatgctagtgctaaaaaagaaactgaatgggct acagtttcaaattatgatgaatttagatttttttaatccaacagcaagagatattatatt tcattcaagtttaggcaattgaaagatttggaaatctttaaaaaatttttattagtattc agtaaatttttcacttatcgatgaagacatacctaaaaaattacttaatgaaacaaaagtc attgaaagagaattagaaaatgaatttttatcaattgtatagtgacactagattaatgatt attaaggaattggaatattcttcagaagatattaataggattgaagcaataaaattatct caaataatattaaatagatttattttccttttgtttgcagaagatttagcgcttatggag gaagagacaactgctgatgtattattaacaccttttaaagcatagaaatttaatcggaaat acaatgtggaataggttaaatgaattatttatttttgcaaatcaaggaaataaacatagg agaatacctgcatttaatggaggattatttgaggatgatttatctaatttgaaaataagg gatgaaattgaagataggtcttttctttgaaaattggaatttaaaagaagatttttgaagac aaatatgaggatattgctaaactaattggagtttataaagatactcttaatccaatttttc	

FIG. 4A-11

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	attaatttattataatttctacatatgacttgattcagaacttgatgtaaatatctta	
	ggacatattttgaaaacagtattagcgatattgaagagttaaaaaatgataatcaagag	
	caaagaaaaaaagatggagtgtattacactccagaatatattacagattatatttgtaga	
	aatacaattattccatatttaagtatttctggtaaagccagcacagttcatgaattatta	
	tacgaatatgaatcatctaattcattggatgttttagattcaaaattaactaacatcaaa	
	gttctagatcctgcctgtggtagtgggagcatgttaaataaatcagtagatatattattt	
	gaaattcatgaagcattacatgcaagtaaatatgctggagattcatctttagatagattt	
	tttgatagtttagaaaaaacgaaaggaaattataagtaatatatttatgggttgatttg	
	aacgaggagtccgttgaaattactaaattatccttatttcctaaaattagcaactactgtt	
	ggacttaaagaaggtttcaacttcctagtttagataaacatattaaatgtgggattca	
	ttagtggatgatgagtcaattgccggaaataaagcatttaactggtatgaatcatttca	
	gaagttttgaaagtggcggatttgatattattgtaggaaatcctccgtatgtggatatt	
	aaagaaatggatgaaaaaacagcaaatacatatttgataattatgaaacttctttaac	
	aggataaatttatatattctacatttgtggaaaaaagttattatttgctaaaaaatgaaggg	
	atttttcatttatcatgcctaattccatttttattaattctactcttaaaatcaga	
	gaactaattcttaataatacatccatattaaatattgtaagaacttctgatgatgtttttt	
	aaagatgctaaagtggaacctatttatttttaatcttcaaaaaaggatatgatgaaggaaat	
	aagactaagatacttataaaaaaaagatgatatggatgaaattccaataaaataattattca	
	gaacatttcttacacaagaaagatggtttgaaaataattcaataattatttttttca	
	gatgatttacttttgatttattaaagaaaattgatggaaataatgaaagattaattgat	
	tattgtgatttttagtttagttaactccttatgataaatataaggtatgagtgaggat	
	attattaaaaatagaaaattccattcaaaaataaaacttgacgatactttaaagaatta	
	ttggatgggtccgatattacaagatataatgtaaaatgggcgaaaaagaatatattaaa	
	tatgggactggttaggagcacctcgggaagaaaaattctttaaaaatcctagaatttta	
	ataagacagatttttatctatagctccttaaggaatctcgaaaaaggatttttgccgcttac	
	acagaagaagaattatataacgctcaaattgctttaatttagtattgaaagaaggcttt	
	gatgataaaaatttattgaaatatttttaggaataattaatctaaaatgatgacttgg	
	tattatgaagaaagatttatggataaaaataaaagaatttgcaaaaatttgattgaa	
	aatgctaaaaatctccagtcattattaattcaaatttttagatgagatagtatctaat	
	gtggattcaataatagagttaaataaagagtttatagtgtaagaaatgcatttcaaaca	
	tggcttaaaatagaatttgaaattgaaaaactctctaagaaactagaaaactattatgat	
	ttaaacttgaagaattcttaaagagataaaaagaaaaagtagtcattagaccaaat	
	caaatacaggacttgtccgaattatttaatgaaagttaggaaaaatagaatatctgcaa	
	agagagattaaagaagcagacgaaaaaattaacctactggtctatgaattatatggttta	

FIG. 4A-12

GAMOLA ORF	Nucleotide sequence / Description	SEQ ID NO:
	aatcatgaagaaatagaaattatagaaaatagctttaatgattaa	
2065	atggacgatgaaacattaatcctaattgaatacatcagaaacgcgccaacaagagaaatggtgcttaagtcatttgaaggagtggactt tataagaccaatccaaatttctaggaaaacaggcattcacccaaacaatgtcagcaagaaattaaaggatcttcgcgaacatgaactgg tctatgtcataaatcctgagtaccatgttcctaaacttttacaggcttactgaaaagggtaaaaatatgcttcaattcttgtag	141
	Phage integration site, genome integration locus	
2069	atgaataaaaaaattatcttatccctcctttttagtatattattagtagctatttctgtcctctgcagttgcagcagcagatgctgatgtcac atatataaacgatgctgcagatgtagacgatgttgcagacgaaaaagttgctcctcttacagctagtgctgatgcacaagacatccaaa ctaagcttgataatgctaaacctggagacacaattgaattagaaaaacaagacatatgacgttgatacaacatttaat	142

FIG. 4B-1

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
	Flanking genome region, genome integration locus	
1980	vrnmknkslilislllllitiisigsvvatdneeinmdninnidnnedianidnvdnvdnsninnptdiridnsnlnreteldsnlnksn qiredeleqsnaksnlkssklsstitvdgsdenqmsnptiqsaidsanagdtiiitgksyvhchfivnkpltiiseigtsmspcpsntk gsgahgifyispeasgtvlkgfnltntygdyddygilirgaenveiinctintvsdgdgirienatntkiadclikdsni	1
	Integration lytic/lysogenic switch	
1981	mktfrqqlledpefqnyllqrpnltesslqsylnaatnfvrftgepfyktvhelrsqqndrienniiirfnpnqsrinimqfefieylk grgctevsidsyvrymrtilstlgiilpkspkldtpqdwylltkddikyvldtanlqykavinfaavtglrvrdmrsltikdfmtate eyhgctevedfldsapdgmigfwelfpqktrkfrlpckvcntpessdlllfslnervkyfewknekdgtdlkitknd	2
1983	mntlkiecskddyiiktakansenelkvevpenwncdyvnavlweedicevlekgdermllipmcgelllegvqedeyikyiclpvkyq dqmvliaki	3
1984	mgmkkvkieskaknnmtrneklfyklyddlynedrlilfstyfniydkfifkkdiihyvlmnyseneiieamkkideiqsegidiktfi pkkycpkckkvmdsygkicpdcgtiliedekkiqelqakdkayeeylekeyniylqnsyhnmiqgsytikirtrkpkteteivripaqt vtsrgkfnsisqhypsetytrqkvtrakykecrvlfdkekmilnidgtatklyydevceleypeqevnelvtltlhn	4
1985	mktqdliniindeespvflnrevfemdyvpdiykyrdeqlakmamycnsipdniapknlqlcggnatgktttlkqffkmlneafpnivt vyincqlfntentvygkiynklygvkgsingksntmlfdkivarlkkenkiliigllddfdsfksrdglnkmlynflriheaeegiqici ftvsnkgeseslllpsrqsstgfrysltstpwsrcttywttgapsvsila	5
	DNA replication	
1988	mflekcdgkdtismslqekmnliletmeskgspfiscihcnipineaeewykngeigdqdfinfyddvnlieesfgfeiykkseyptlh tqsinqiastypmnrtqnektplflrrerklyeitnifkshstneifismdskskiklkheikydftlkelneifknyleedcsiyilndn rafvmtlghfqfefdvfgsakesyvfdveiddekydkifirssyyfnipvdeldglakilkqkriieegffegqfly	6
1989	miiflklkfgrvimeklieidgvqyteaqirralaierdvrspnfvdmllgkikpselasrvsekgda	7
1990	mcyvgntrtlvyhtedcfcnhwllnenktileekpvdmkpcsfckpqfdte	8
1991	mldmvaemvenirkgegdgysiyppfscivflqgkkyseccckaeardqkfalvnlvgfrredvkvidprtneelfv	9
1992	msmladfeparlhkrtwaerhdveilaviclaisiamlllffalaeptvagvi	10
1993	mtkefedfmrrntgllvflrwdtvaylkyleshyderkyecayrlleaidnlfdfyqitfskktereipdqlfekdkinkgflshigka ckkqsdfygqrwkstrgirdayghysagnflfadiygclhrisddcyrilnfseyeieeka	11
1995	mneivtttnennvpvdvdyaieewkayqrltrelldetdyqthrgrkyktksawqkyarafnintqiidkeivkndkgivieaeytvra tlpngrfvesdgscdrresgkremsnhsikataktratnraiseligagdvsadeldpafdkvqhsktnhvieaevaeiiespydknag fetadkiepvdedpvcknwvkticktikaegkpclkgvliqkartigtmtdeernrlieyiktlpkgevnldd	12

FIG. 4B-2

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
1996	mieciqeegdftdwevpssssdikyivsvddegnlfcscpdfyyrksrmnphisnpesyckhirqvleednrlqml	13
1998	mdninktktslakfeeffstvykdevmevlekypeertlvvdyenlemfdpdladlliekpdeviaasqkaiknidplmkdpkldikfknvsncidfvnadskyigklisfeakvmeakepkpildiavyecrgcmslreipqtinsslepslcpecggrsfrllqdeseflesqllivssddtskslkvlllrdecsfdlysmgqevritgilksfssnygyeyflecnlieilndsedseydeygnrnspeyr	14
1999	mankirvnltvdpnlwqlakdklpcsrseffenqlkmflgieddeseiikdiqtkeneinalrdklchvrkskqlklesnksmekamaslnrmhkkygkigenqirnlahvhkvdfddlkkecqdncmnifefaevpkhdsvm	15
2000	mtigsldnfgksdgenmnpedfdcsvffemykalfeildvevgsfaelldvyknvemdytlkrhalkqkeilywfntdwkeelgkekptekdkekwirqkigydsfvveqlevklkhirrmyetalkhsfeaik	16
2001	mnvktvmndliglskefegveyeiesknsiyfysfpkymkegivilkysaiydlhtilkgmdgiivdilevednpgdekkdllyvqievke	17
2002	misdeweeeyyvkvneeleqvevrffskvdrlvfaqsyssssfsfeeaeilcdricdiltndlgnakyylggkdel	18
2003	mnyrewvasqfelrmtedgrvcfktpcsyytfskedfeiiremflnfes	19
2004	mdrikelnvcgtckhshlipdingeiavnicrigseavnkdggltycvdwtprrkls	20
2005	mlskkeaiqmtldnekhypvkckycgkpftkshnrqmycsdscrrnalreqkaryqakrrlkikqkvlivdeykkyglgsygtsanghrknnfsheymaiqkemkriglkr	21
2006	mylakfcpncgnkveendkfciycgnklrviipekkvkrssnsindektskyvevidglmrykvfpsslpvkyiiykvnygttsdeiknilengnynykinihyflqnkklyfrspknpnmffkfhndrfneemrkdnkeiihissyarvhrpsltkifnfnqektfilankhfeenirktrl	22
2007	miydkatvtslivaillplmsmlgigeltqnyilaivsgmialvvwyynekhnsdlvsgttkcdcelcyggddeali	23
2008	mkhlyeiipyrrtvwitgflkttvssamittgvvilfnsitehpyfmewdeigivlgivsitiaciyiamidrwkerrkkeeldtiedyinrkaeeianmkvlrkleeleee	24
2010	mieistikitdikpaeynprimsqlehtklrnsmetfgvvdpiiinlknnhiigghqryevlldksmednefikelhlirlgdvgwafpesdleveddhekalnlalnniegewdlpklepiltdlkdvgfdieltgfsdieltelnlennlvfaeefepdeseedvdlediydepvkemlqcpacdhvdvvkrfkrvdsqgd	25
2012	mdvdnllgydkvlelrslleevtdk vipvwhknrgikdfkqmcqdynfvsisgwrnedvkddqfihfvrhahrngcrihglgltrrkvldrvpfdsvdssswlqtilyarlgqkqldskfaterrgdlavlsyikwmktqeeyykkwrhyhd	26
2013	minerirpylnfsfndkkvifvtlfvvsnlisnllaikvfnlgfwglttdcgnllfplgylmadvitevygertarrvillglfanillivattltvympypsywtgqgayaymfgftprivlagfiaylvgqfvnarlmvlikkwtnskylfmrtigstlggelcdscicssiayygivpnsgillfilmqyvvkvtwevvmqpltyksiawarkdg	27
	DNA packaging	
2014	mpepwerqrdengklepikafeyfteyltmdkprsmrvlcerlgkkdgyirqlhaysstwnwveraeaydehiilkkrlrkekfydelv	28

FIG. 4B-3

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
	eselpnlrkrleyynknmndietdmttkptskahaydknskahsttlneillmigkpteiketsleadiesdnkidlenrvevditsde fmeseleymrkmieek	
2016	mkdivnhygylspykptirsdskaknkfklnepyrgqmlsagaggsimgygagllivddpiknvaeaeskvrqaklkdwwggtiksrvq rrsnglppikiviaqrlhlkdlhgiiketeptipandafrilrnggsidpntwvdfnlpaicdseddilgrkigevlweeqrdyewlma ekrsmgsylfnsiyqgqpverdgeifkrewfqdevnhkltclidpkdipkdlpqlrywdfgasgdagdgtsailtsy	29
2017	MSKKQEmmrierlkhyayqtgliipifknhdlikkiengkitntdeikiyieenekqikrrrefisiiydnckyfkydsicykliskvn nfeiksleelmneiesekkkngfrknieekelikenenkgrlkeyvrivrreygldftsvkklklkidnnkirskeelneeiieekkkt elrrivydsdlnwdlkweleskirhneittkeelikeirriefiniiydnnrdfkldyvtiqklvskigsneittkeelikem	30
2018	msknakadafvvttedgsydivdadvleryaiksesdetgskqlktdgweyddtlleplydplqlcelleintyhencvdvvardsagi gydivpvtgekekelnkpkltnfleniepninellyqmnydrratgygalelirkdksksepvnlshissytlrrtsdgkrvkqrvgtk tvwfviygknydkegnlcdvhsetgefhpynslskeeranellwtmeyttkskyyglpkivgaipaiysdisrskyn	31
	Phage capsid module	
2019	mmalkvrhskrqiqnikreyrrrlileeqcsreiadffrrlerkihkvmdehweselglfhlnkvsdiiqdsrqeyydilfkyckdsym kgreaterrfnrklenismkadvnitrledlfkpdptirynlnnkvfqasahtmdrvdnrimenitqsyddglgideakdrltveyngl ksweaqriarteinsaqndgafdvygelgveyhqwwtaqdervretpqadhrelhgkivkvgnsfsnglqypgdrt	32
2021	miieipgndrteeinlpngqfvlitylqendmmslpdgkyvcpfrmiqlfteeggeligecieenphyntrfyntvfehldegieyr	33
2022	mlrvvertyyqqeeiktldcrireagvntyslarqgavdyptyesynevreerikeakekygesyyyhwrdvetyfyylgrffsdleei ekylertvtykphreelkeamerldkrfeevinefwysleeyeditedvleqlkegdctvhceflikefkkfvnvicriikqneinhke f	34
2023	mdainvinqnkilvdvlyrgtvnlidivigdalvydnptvvvkcyttdlafatteideivleneefellityedgefsillkshnlgel gyiewvi	35
2024	mvfvdlceneiadmvesfyrngdgssriltnravgeivehycsfevdgrittnlrdfllysvvlydtigeavddgvnleevfiiedrnc ytgksqviligg	36
2025	mvnwlkaigdnfsvdylllalfssgdlilvaivlnsygvispenvrelvidyisyrkvdifwrhlrrprmsfedyvldnfeemetgelt reqvvefvsrqerkgltfcneifiavplkkgskddiveilwneyfvedykenwleqhenlgwndwkkllkkeivenggddfqifrnhli dcvlmey	37
2026	mtmkvtfedengertvefgddvdfvliesdddgnieiregdweldgdsdddweeyddwdeeef	38
	mdefvetlfdtywkvnengeymsltdcgdfyiakvapcvrnwsiviecncfcfhckefvyhengaileigmeisslylsqmeikdlkiy midc	39
2028	mflegifeqdgenvreqviywrksnqvhnwfvvnaqdgedncqphsvsreqleelrdlcravladndkaeellptrpgfffgaidydew yyydlqytvekidevlkddrylyfeycsww	40
2029	mqlvvegenmecpcdncemlrenepikvidwksnspfgngixihfwrerlqdvgyng	41

FIG. 4B-4

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
2031	melmtremegklksfpfysqdgkgddaivvmkffnpyglgtwyvleaekqengdylffgyvespitpefneygyfslselenlkipiki ngitvsygrierdlyfervrigdiign	42
2032	mfgqkkefvkkmyhvgdvvelvhmddaqappsgtrgeilfvddigqihvrwengsglaliygedrfkvverkge	43
2033	mdllylyddltarrevydsvglsfvvkykfssrkeaqdfalkygaeliee	44
2034	mierrrlgmkydiftildeisrklddgelsdeqvdfllqmeilveegtitdeqaqdvmngdy	45
2035	mmklslkelgeeiemilaeggltfdqvdyllyletciadgsiteeqkreiicrdf	46
2036	mvreqerlimyleidevkcenidriefddvameivltdekvyerikrwlksneidydcredryfanlieyviritww	47
2037	mvtvrdvgftieerffltaqeleysevgeehesvidraiallytklrtrdfeftdeerelledafvivsdq	48
2038	makenvidykierqndntwsylyvtergrgniiassfgelrqkvlkrglpwndisnlftkkssdsnvrresvkniddesvladvakkss dsnvrleavrkisdnyvlidivknasdydvrreavrkindssvlediaknnndenvrleavrnindesvlenisknasdskvrieaikk indetiiiklaknnndedvrieavrkindktviidfaknasdskvrreavrkindssvlayvlkndpswivrieavr	49
2039	mvvkcpncfsprvskcedtnikwqcdkckckfnhgafdinaemekveqltiekiererererteqferaikeakeqferererteqfererke rlerekrekerekiekererkerlernrikieerererikrnerirrenernriksdkrerekseekrikrnerirkanernsikrekr ererernvmtideyyrsigygstgkskvwsaiiipillviciiililmfygggm	50
2040	mtegqirqiaheylanyslvdknheffetrevigvpvesyitnepislkgldgtvneypkgtwiattritdeeemekalngeytgysit tvskkfadkqiqlprrvlmkdikdpvgftislvrkpcvrgakfcsmkediengdvvseniddkleeetkgfvqsikgifnkedkdedkn pededieldikaivdevtkdfvntddfetfknelekalsdkfetlgaelfkslkkslekdkaeeakksgr	51
2041	mgneatlnqlvneqekavfksmrtdmetgkavlnveqlgyflreatldntilrdadfklmksfkkhlnrvgingrvltngydvngetdp eipaadvdfganeldvkklkamceieddekednmtqaqfeqtllqmmgerigedleywalfadsevarsddpllntndgwlkkcanhis srsiapsngmfdiedgpeamfdamikalpprfrknrrmlkfyvpfevedayrnilinrgtglgdsaqigfnalsykgipiehcstldde dgrgmlgnrvcsm	52
2042	mrrrclnspehngmishiivllicfiglveailmalvdwedlaisvrksprklynvlkdelglpewnelsvierrsmkkryavirdsfp elppweelsvidrrshkrlykliksvydgdyddspslegppaavgpqkeipleeaeyp	53
2043	mtwigtedvieftgvkpqtfrfekgdtssletllekwilqaegliisycnydfndleeippavvnvclrltanmvalaqarkdtpviqv kewnvqtvssnifsndlkrdltpfvherksykgdeidffvitgdddsw	54
2044	mVklqidveelkpleprfkkvakrtvvltanelqrnlkklspvdhgrlqgswvifqtgelertvkssakyaifvndgtglygplghkir pkngkflaftpnkgkfkgklvvvpwtrgqkpqrfversmemterrvqefmirammemds	55
2045	mrfvntaslvpqtvkaylereicegglledvetlipsvnsdvpvdppaiwivqhpttrwsgsqpnlsnkiamsvpfefvcveysddlee aeilgislasrvgsslmknfnkvkvddsmpnrffhklefetlypvgevtvvgkseripatsiifnfvyvvdwlkcnrryd	56
2046	mgirvvgmkeearygvaesapdfhqevskakaslnstpntkssgsrmkkkaragvykptaniegevdlkrighylkafldnyhftdggs npnvhefwggennklssftlwvtfdifektivgslldnfkmevsdeymkftadfvykteesdeienielykvklldgdwalmfydvsve idenappgivssfsfdgknninvdktiglgsrgpqrkaaaqgrdisisfvstleretleliqkaeygevgtepseck	57

FIG. 4B-5

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
2047	mvvvkksdilkgvkkiekvkiealdgdemylrplsqaeinevdeieakamgifetnetahrgrrqkpksvveskgkinlelqqkaqhqa ktkaiflsldneknvgeeawseteieqmphklfeelfnhvkrlsgieldeddvdtfh	58
2049	mpssnvmniivkaedmassvaqkvensfrklgntidstfttslsntkfnqeltsfgtdldkvtqrlkqvgvngqssfnqltnaerrtle klsefdpvsaqvlqhlsrigitgqqtfnqlsvseqkslmnmkstaqqleevnsklrvvgigaaqaanmlnqmkldpsvgsnldraklkv semgysldstkgkilvlgtaiqtslgnkwdsiktkvqttatnirtalgnaltsvkskvqnlgnafsglggiissaig	59
2052	masvtkypsnvsqttggkfvsfsnlaniknnadgahavssvlikskkqspnrpstvsckgfgfslpegaeptkitvtyrhrknagsdys sknkthicniggptisllgvsgfsskgsgctttmtthtkafsvhgklsraqvnsanfgvkldyptnsntyngymrisyvrvtveyitsq ysvsvkhvsgtyededyvvslgisnknltsynptctltvpagftykgvtgaatgtvtkvnnrtfswnpqlqgragsrqislafepnvtf pegtdsy	60
2053	mgiaivvmdneenflqfldpdlctinetieelglrtlefnykfqdyvedrdlfrignkiwisnsqsledclyvintpvensvyqenyfa ceieevlaelyyaplfsqteltsangftlrttngeqtvdvdwnalnywfglyfnigvvqeclgtyanritvngtmnrlnllrsieeqtg nrfvtryekdlldntihryldflnpvnvsknwklnieydfiyeddgeyceaytsdgnpiseiyddieeeddivdfpp	61
	Lysis	
2055	mvekitvspqevrgygnvvdekeledygsyrcdvsessevikgveerifsvsgvpapalsianvtedtrrgrcahisasfedgegdgld dkaislksgddvlatittgsgenvfdvvlydsaqlyavfdgddyyppavseaitvnpakslwdvefildeeeyevgdtailsgtvgtiv deivdgeivtrrqmeanvtltlvtdlgirrcstnangefvlqvpniqqnqwrvviaatsthlvfnglidvpvhdysl	62
2058	mvrfsrdmlqdgakrmfkwlrkgeglpnylimydmdrnkeyklvpkeyaglyesrnifwikngrepnyvtltsvarnplvmdyqntnyt ccptslslasqmlyhyksesecakalgtskgsgtspaqlianapklgfkiipikrdskevkkylkkgfpvichwqvnqsrnckgdytgn fghygliwdmtsthyvvadpakgvnrkykfscldnankgyrqnyyvvcpa	63
	Lysogenic conversion	
2061	mkkhcfyflgdsfadicneamfcekhlvegnyldsiiragkaseiitvnicelegqdglissgqkkrlemlgykgiisydiykrlnhir kirnkavhghlsdiednanilhaylylicayfykeyrdtnfsaedytgpimdiaskpketasetsednenigefisspldylfekydd syllnelsklkdsskeaveddnlsefkeylhidrsiqedflkalnratsfnsshlimlcgsvgdgkshlianlkkknpelfnqfaihyd atesfdpeknaidtlasvlepfndnnlnnsteklilainlgvlnnflessyanedytklkliieeanifesnevsdniygdkvsfvtfs dynmfelnddensnytsskyisslfnkitqkedtnpfyvaylkdkdshfinpiiynyemlmdeevqktiidylikifikyrkiistrdl lnfiyeiivppeflksedldnindfmdyslpnllfgypersdllklcneldptlhrnesldkfiidlninddtekilnryfdftrfnfl eeygeylvdfrefnnsekekvtnilirfavfygksiiknnfkdkvylnylkylyayntqshkdykylftevkdaifnwkgsykkntici dtldsfkvyknlklkpsvdkfekslldglflgnrfktdikiyfsvesnkkkiplnvdfslyqyimklyngfkpnqsdkddliildefin nlldedtdddlyvisletyeeflfesndfgtfefkrg	64
2062	mdfsenynillkqmtcdvnkrklihqinqnspllpfktntpkkanfengfdiilgelsrillnktieknfkldnivsnlidnnieiedg tkeyitkllneylfdekndlkishpnlylyiplsnnkssngeqevalflrdifcknnqnlinffesydsnhiilnlilkntpnlhhkit	65

FIG. 4B-6

GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
	etkyvihfeeianlfnedinyailykkffmenignifayyyffyisqlilkiskgfndnnefeklyylldwesasknrkslnsysllkh hskplyakmavidqintllqtnnllekdiseyfnnldinsknnflhflkkwvsdyryvrnfddkelpdnlleltеilfeslkneklgvd gavqsryalnlediakkyllkrrgsygyvlninrdmllvltalcvkdkkiklnqlfiefekrgvyfdkyskeevvnfltklnlidkksd sgdaqyvkpvl	
2063	mlnqfydylsnkllnyfddtkilsgekffisfdeddqimsfynslrsiaetnfscsefiyvhtisgkeyntysiningvkfvisesltі nvdflvtlrnqvtsqegvwkdtallvicneaidsigkgmrnlqkegmplnvksisknledeindsqilnysdkqiakfslniqeeelfq ttlwdyetilsiinkgfvsdedlrelnlfkdddqlnqnspqkmlkrlkenydtfnevnkfsqygdkkeqlknmftdsgvsilskddwyka ewkmvkkskddfinqqnplnynenlekitenglnyweipnsftktgkrkrnnivfnpnhshevslkfsfdqilsnsflntnskkftiar gkslivnftldssepifktikykhkneniseftfnivvlnfepeifnsiksrfsvnvkskqiivtndedsfdivfgtgskeiekliken geklylyddesliiseqspawndgklsfklykdnnyapflikekskktlpvnsyviwnlkrrnmenfifngvkavqdvnsfylveefke flkmereiikqdifyakrnidgslekievsfsneletaymdifnyyktfddspednlpslvylnddlkelykkfitifnkeiseieens ilsdfkykknllklgrietdnkimysplspliayqlevskqcgnedlsvnilerlvpnnlipyicsddgkelfrpiygeeahewliye kseevsigttnvfisnvvteklnqfvkhfnylfsfnnsspikinlinikddkevvkgvfnfirsrlpdktktkkvipveiniyndaeks sfdnlfdcqseiqlleefgikklksdifdpididihmiqnnisyykhpfkkeyeyahlsfykvkshnniandnmdkietglslngllss vtstthkseyrtgfgtnnilnmsnpliktvinlnelvensknfgkntysknksvittveleednieelydkshwvtfieptfgieyfds sdsnliihysdqyssskydtitvtnkstqyeeiirdfdlqskyvkvktdeelydvikmfnsingewllrvisnsghydreklsiisaik yclsildhkdivwipvsmeeilriagnvkldknkgifdsklikgnhsddllfigvkfnednrievifypievkiglnnastikkgksql dntyklltqlqninvensefrnkffrnffiqilllsneqklvtnhiwdekgldrieefkaellndeydilygleeyigkgslvsfkkes hhisiymdvdkqvielpedfayyglatpireihdeiqsdntdilaetllshvdiseiraknndicdsnedmsidddfdnlsefedsfie eeseiseepdeeltdatsssdnesienigespskisnvraligtqkgynhkvywefghpslanrhmliggksgqgktyfiqrmlkelsi qgipsiiidytdgfkpsqlepnfkdslgdkisqyfvvkenfpinpfkrntimidkdifieednstiasrfksiinsvyglgiqqsntly qtvldcldkyddnfdlnilkeeilkdesnsaqtvlnklnellddknpfastdfdwsvldnkdgkvyiiqltalskdiqtiitefilwdlw nykltngsednpfivvldeahnldfsndspcskilkegrkfgwsgwfatqsvkgsmkideiaklenadekiyfhptdvstiakdlskdn edkkiyekelsqltkgycivqggsaidssgnlyqpnpvtvkieeisfden	66
2064	meskdrientifndkrmdayidkyfeeftlsseqeealniwidklnndqltsekgnyhnffeiiiledllgykrsdvkheenigdeghpv efvlekdgkdyviielkgttykdltkrrpggqspvegatnyasakketewatvsnydefrffnptardnyisfkfrqlkdleifkkfll vfskfslidedipkkllnetkviereleneфyqlysdtrlmiikeleyssedinrieaiklsqiilnrfiflcfaedlalmeeettadv lltplkhrnligntmwnrlnelfifanqgnkhrripafngglfeddlsnlkirdeiedrsffenwnlkedfedkyediakligvykdtl npifinlliistydfdseldvnilghifensisdieelkndngeqrkkdgvyytpeyitdyicrntiipylsisgkastvhellyeyes snsldvldskltnikvldpacgsgsmlnksvdilfeihealhaskyagdsslдrffdslekrkeiisnniygvdlneesveitklslfl klattvglkegfqlpsldkhikcgdslvddesiagnkafnwyesfsevfesggfdiivgnppyvdikemdektakyifdnyetsfnrin lystfveksyyllknegifsfimpnsilfnstyskirelilnntsilnivrtsddvfkdakvepiilifkkgydegnktkilikkddmd eipinnysehfftqerwfennsiinifsddftfdllkkidgnnerlidycdfslgltpydkykgmsediiknrkfhskiklddtfkell dgsditrynvkwgekeyikygdwlgapreekffknprilirqilsiapkesrkrifaayteeelynaqiafnlvlkegfddknllkyfl	67

FIG. 4B-7

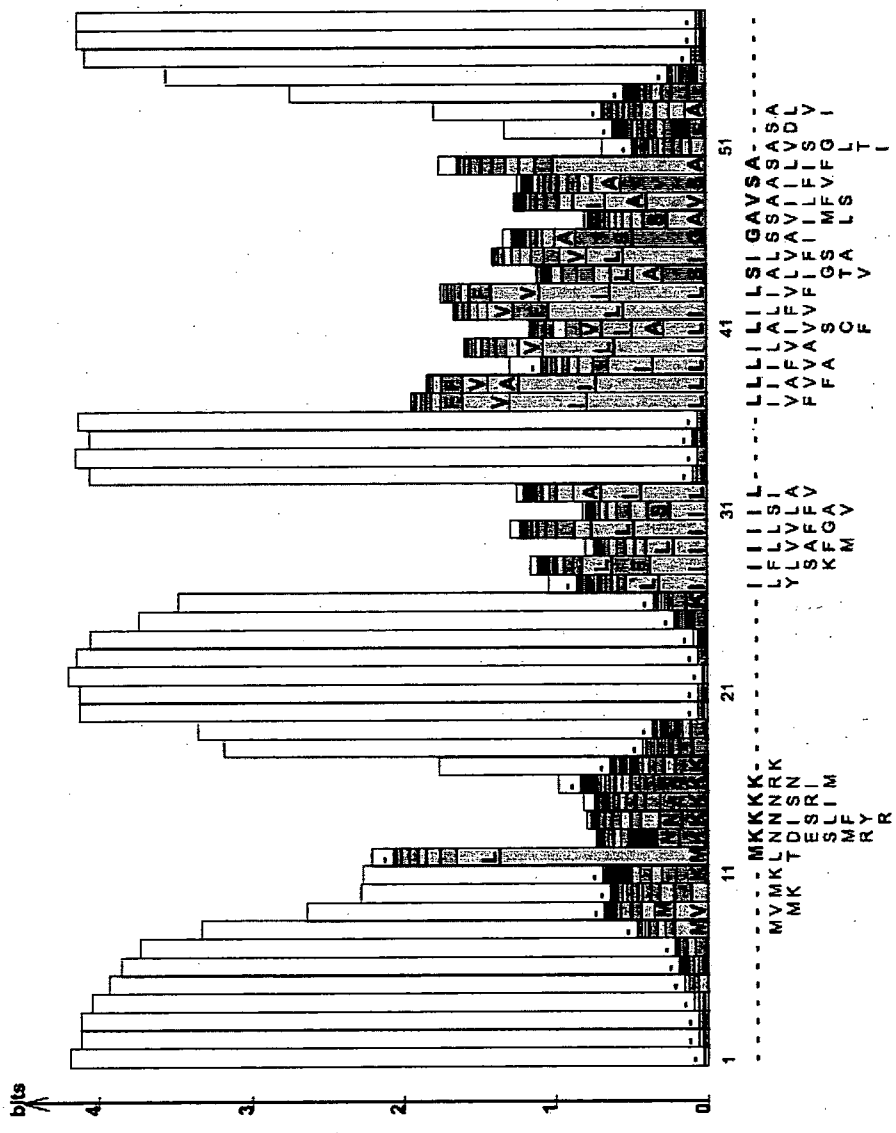
GAMOLA ORF	Amino acid sequence / Description	SEQ ID NO:
	giinskmmtwyyeerfmdknkknfakilienaknlpviinsnfldeivsnvdsiielnkefysvrnafqtwlkiefeiekiskkklenyy dlnfeeflkeikkkkvvirpnqiqdlselfneslgkieylqreikeadekinllvyelyglnheeieiiiensfnd	
2065	mddetliiieyirnaptremvlksfegvdfirpiqisrktgihpnnvskklkdlrehelvyvinpeyhvpklyrltekgknmlqfl	68
	Phage integration site, genome integration locus	
2069	mnkkiilsllllvllvaisvsavaaadadvtyindaadvddvadekvaplasadaqdiqtkldnakpgdtielenktydvdttfnvtkq vtikggdtvikasgasgggsgalfianeagtafegitfintdghknygeqvsgyaiqlaiengtvdnckfidwssgvygkgasfcsitn syfngssegvtnggkkeygtkainlmgshditvtgctfeggvldaisiasnsgnnimtdntfidncyaiyfggastq	69

FIG. 5

SEQ ID NO:70
SEQ ID NO:71
SEQ ID NO:72
SEQ ID NO:73

```
                (101) 101        110        120        130        140        150
          PeiP  (98) SRYNRFKEVNGREPRVVFLYSG--GGPSVSLETFKDMCKRYNQFLEENRR
          PeiW  (83) RRYEDFVRINGREPNLIYLEQG--KSDHVSLGTFKDMLRRYKDFVRINGR
MRU ORF2058  (74) ---------VARNPLVMDYQNTNYTCCPTSLSLASQMLYHYKSES-----
     Consensus (101)  RY   F   INGREP VIFI NG       VSL  TFKDMLKRYK FL   N R
                                                                  Section 4
                (151) 151        160        170        180        190        200
          PeiP (146) EPRIVYVTPPEPPVPEEVREMRRVLGEFK-TATQLYTLVSRRCKYKFYYN
          PeiW (131) EPNYISIQPQPSLKGHWTTKVIEKIGTFH-DATSLYERVKKTCKYKYYYN
MRU ORF2058 (110) --------------------ECAKALGTSKGSGTSPAQLIANAPKLGFKII
     Consensus (151) EP   I I P           EM K LGTFK SATSLY LVAK CKYKFYYN
                                                                  Section 5
                (201) 201        210        220        230        240        250
          PeiP (195) DQTPNREALKKMVTDGINCTDACQLFKPVIEGLGYSVRIEHVKVRCNDNK
          PeiW (180) DQVPNHVAVMRMTTSGINCTDACQLFSKVLEEMGYEVKIEHVRVKCNDGK
MRU ORF2058 (141) PIKRDSKEVKKYLKKGFPVICHWQVNQSRNCKGDYIGNF----------
     Consensus (201) DQ PN  AVKKMLT GINCTDACQLF  VIE LGYSVKIEHVKVKCND K
                                                                  Section 6
                (251) 251        260        270        280        290        300
          PeiP (245) WYGHYFLRVAGKELASVSLPSERWTVWDYVSATKTGRPLGAPCCSRGIQH
          PeiW (230) WYGHYLLRVGGFELKDG-------TIWDYVSATKTGRPLGVPCCTAGFQH
MRU ORF2058 (180) --GHYGLIWDMTSTHYV--------VADPAKGVNRKYKFSCLDNANKGYR
     Consensus (251) WYGHY LRVAG EL  V        TVWDYVSATKTGRPLG PCCS G  QH
                                                                  Section 7
                (301) 301        312
          PeiP (295) LGWGIVSPKHD-
          PeiW (273) LGWGIVGPVYDK
MRU ORF2058 (220) QNYYVVCPA---
     Consensus (301) LGWGIV P HD
```

FIG. 6



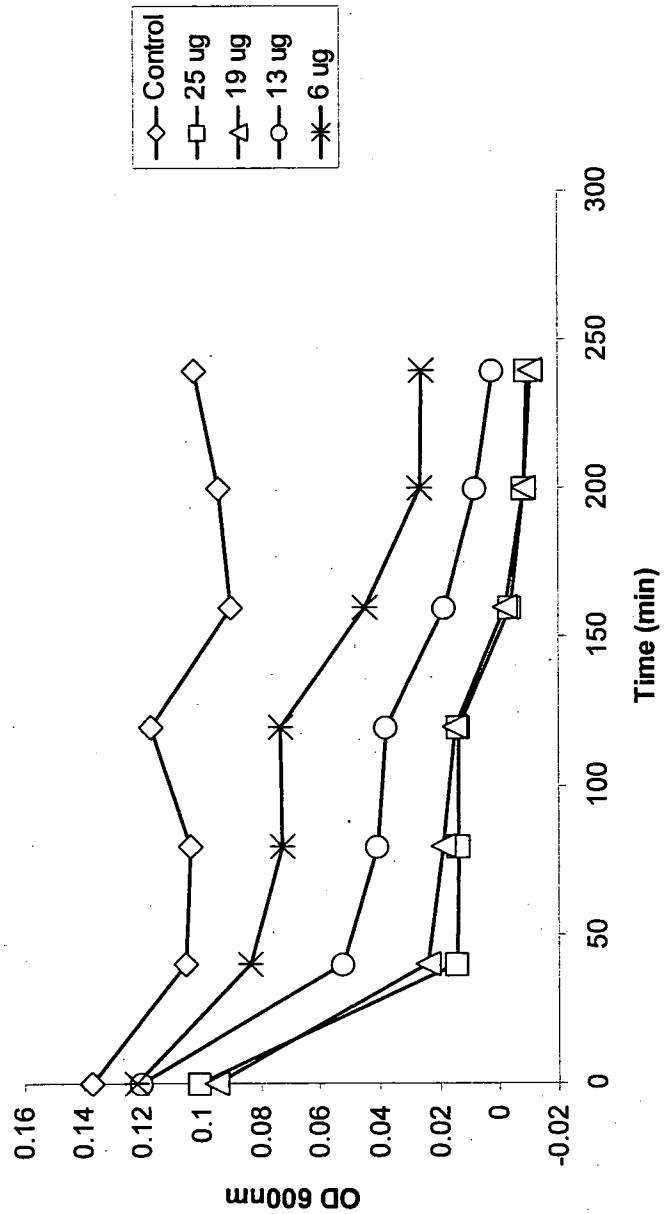


Figure 7. Effect of ORF 2058 lytic enzyme on resting *M. ruminantium* cells

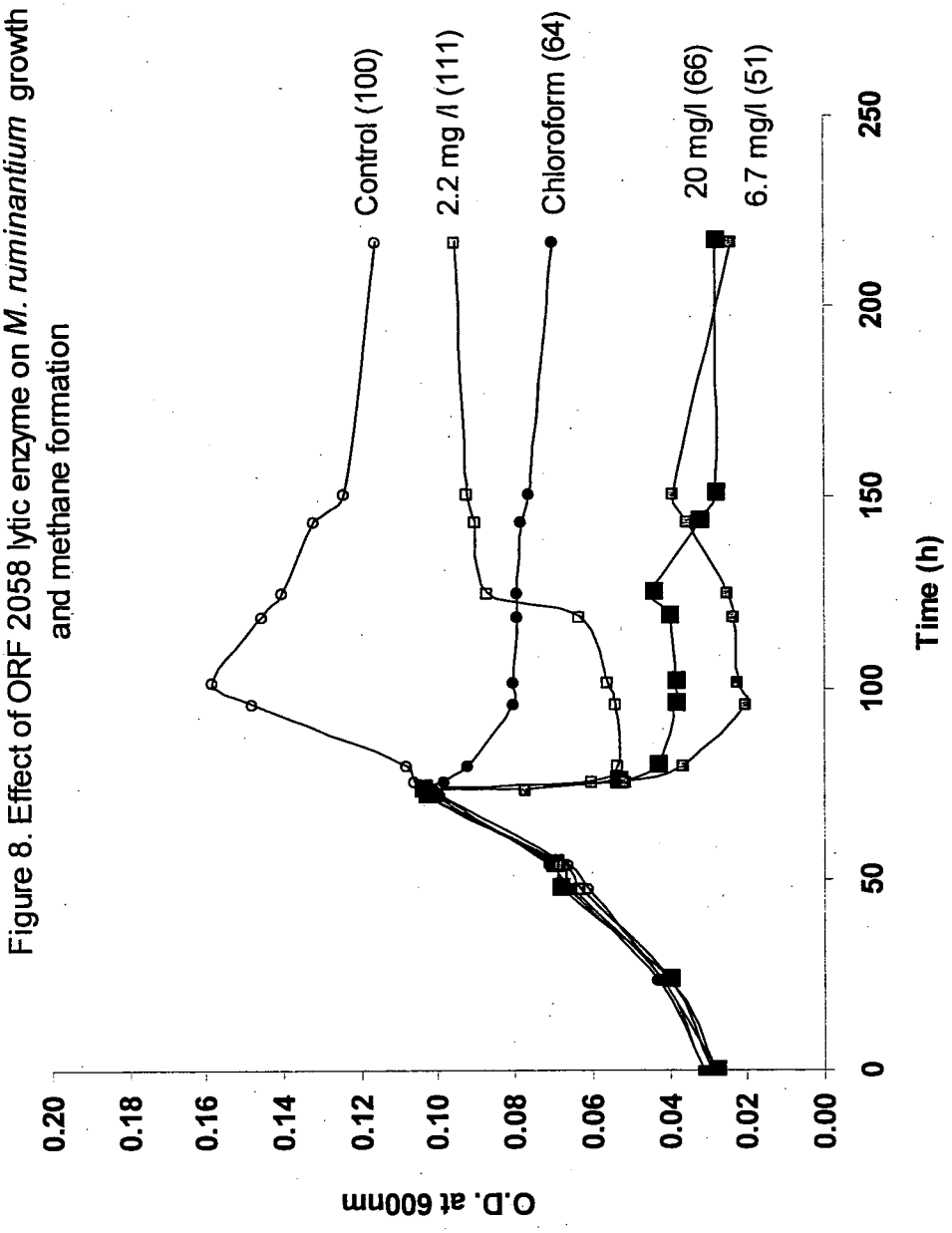


Figure 8. Effect of ORF 2058 lytic enzyme on *M. ruminantium* growth and methane formation

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NZ2008/000248

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

C07K 14/01 (2006.01) *C07H 21/04* (2006.01) *C12N 7/01* (2006.01) *A61K 38/16* (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
GenomeQuest blast search in nucleotide and protein databases based on sequences 1-142
STN files CA, MEDLINE, WPIDS, AGRICOLA, FSTA, FROSTI (keywords: methanogen, methane, phage)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GenBank Accession No. CB993593, Homo sapiens cDNA clone, 1 May 2003, NIH-MGC. 83.57% match of part of sequence 80	20
A	GenBank Accession No. CB991989, Homo sapiens cDNA clone, 1 May 2003, NIH-MGC. 83.57% match of part of sequence 80	20
A	US 5985907 A (MEYER J. WOLIN ET AL.) 16 November 1999	1-57
A	NAGLE, DAVID P., JR, Development of genetic systems in methanogenic archaebacteria, Developments in Industrial Microbiology Series (1989), Vol 30, pages 43-51	1-57

☐ Further documents are listed in the continuation of Box C ☒ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 November 2008	21 NOV 2008
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. +61 2 6283 7999	Authorized officer **Sonita Singh** AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No : 61262832658

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/NZ2008/000248

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report	Patent Family Member
US 5985907 A	NO FAMILY MEMBERS

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.

END OF ANNEX