



US008586826B2

(12) **United States Patent**  
**Mori et al.**

(10) **Patent No.:** **US 8,586,826 B2**  
(45) **Date of Patent:** **Nov. 19, 2013**

(54) **VIRUS VECTOR AND USE THEREOF**

(75) Inventors: **Masashi Mori**, Ishikawa (JP); **Koji Dohi**, Habikino (JP)

(73) Assignee: **Japan Science and Technology Agency**, Saitama (JP)

(\* ) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 516 days.

(21) Appl. No.: **12/450,999**

(22) PCT Filed: **Apr. 11, 2008**

(86) PCT No.: **PCT/JP2008/057170**  
§ 371 (c)(1),  
(2), (4) Date: **Dec. 14, 2010**

(87) PCT Pub. No.: **WO2008/136253**  
PCT Pub. Date: **Nov. 13, 2008**

(65) **Prior Publication Data**  
US 2011/0107466 A1 May 5, 2011

(30) **Foreign Application Priority Data**  
Apr. 27, 2007 (JP) ..... 2007-120181

(51) **Int. Cl.**  
**C07K 14/005** (2006.01)  
**C12N 15/67** (2006.01)  
**C12N 15/40** (2006.01)

(52) **U.S. Cl.**  
USPC ..... **800/280**; 435/5; 435/69.7; 435/6.1;  
435/475; 435/320.1; 536/23.72

(58) **Field of Classification Search**  
None  
See application file for complete search history.

(56) **References Cited**  
U.S. PATENT DOCUMENTS

2003/0036641 A1 \* 2/2003 Padgett et al. .... 536/23.1  
2009/0017490 A1 1/2009 Mori et al.

FOREIGN PATENT DOCUMENTS  
WO WO 0206501 A2 \* 1/2002  
WO WO 2005/049839 6/2005  
WO WO 2006/003018 1/2006

OTHER PUBLICATIONS  
Marillonnet et al. (Nature Biotechnology, vol. 23 No. 6 (2005) p. 718-723)\*

T. Gillespie et al., "Functional Analysis of a DNA-Shuffled Movement Protein . . .", The Plant Cell, 2002, vol. 14, No. 6, pp. 1207-1222.

R.L. Toth et al., "Improvement of the movement and host range . . .", The Plant Journal, 2002, vol. 30, No. 5, pp. 593-600.

K. Dohi et al., "Inducible Virus-Mediated Expression . . .", Archives of Virology, 2006, vol. 151, pp. 1075-1084.

S. Marillonnet et al., "In planta engineering of viral RNA replicons . . .", PNAS, 2004, vol. 101, No. 18, pp. 6852-6857.

I.E. Johansen, "Intron insertion facilitates amplification of cloned . . .", Proc. Natl. Acad. Sci., USA, 1996, vol. 93, pp. 12400-12405.

S.J. Yang et al., "Construction of full-length . . .", Archives of Virology, 1998, vol. 143, pp. 2443-2451.

Search Report dated Sep. 23, 2010 for corresponding European Patent Application No. 08740267.3.

Koji Dohi, et al., "Insertion in the coding region of the movement protein improves stability of the plasmid encoding a tomato mosaic virus-based expression vector", Archives of Virology; Official Journal of the Virology Division of the International Union of Microbiological Societies, Springer-Verlag, VI, vol. 153, No. 9, Jul. 25, 2008, pp. 1667-1675, XP019634121 ISSN: 1432-8798, the whole document.

T. Satyanarayana, et al., "Frameshift mutations in infectious cDNA clones of *Citrus tristeza* virus: a strategy to minimize the toxicity of viral sequences to *Escherichia coli*", Virology, Academic Press, Orlando, US LNKD-DOI: 10.1016/S0042-6822 (03)00387-8, vol. 313, No. 2, Sep. 1, 2003, pp. 481-491, XP004452427 ISSN: 0042-6822, the whole document.

J. Yamaya, et al., "Expression of Tobacco Mosaic Virus RNA in Transgenic Plants" Molecular and General Genetics, Springer Verlag, Berlin, DE LNKD-DOI:10.1007/BF00425710, vol. 211, No. 3, Jan. 1, 1988, pp. 520-525, XP008024750 ISSN: 0026-8925, abstract.

\* cited by examiner

Primary Examiner — David T Fox

Assistant Examiner — Jared Shapiro

(74) Attorney, Agent, or Firm — Harness, Dickey & Pierce, P.L.C.

(57) **ABSTRACT**

It is intended to provide a polynucleotide comprising a viral base sequence, the viral base sequence containing: a first base sequence encoding a viral replication protein, and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition. By using this, a vector containing a viral base sequence is constructed, and a protein is efficiently produced without worsening growth of a host cell containing the vector.

**21 Claims, 5 Drawing Sheets**

FIG. 1

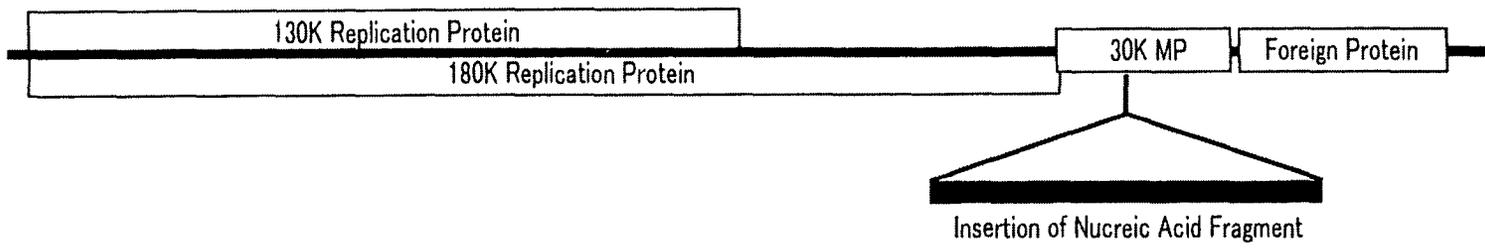


FIG. 2A

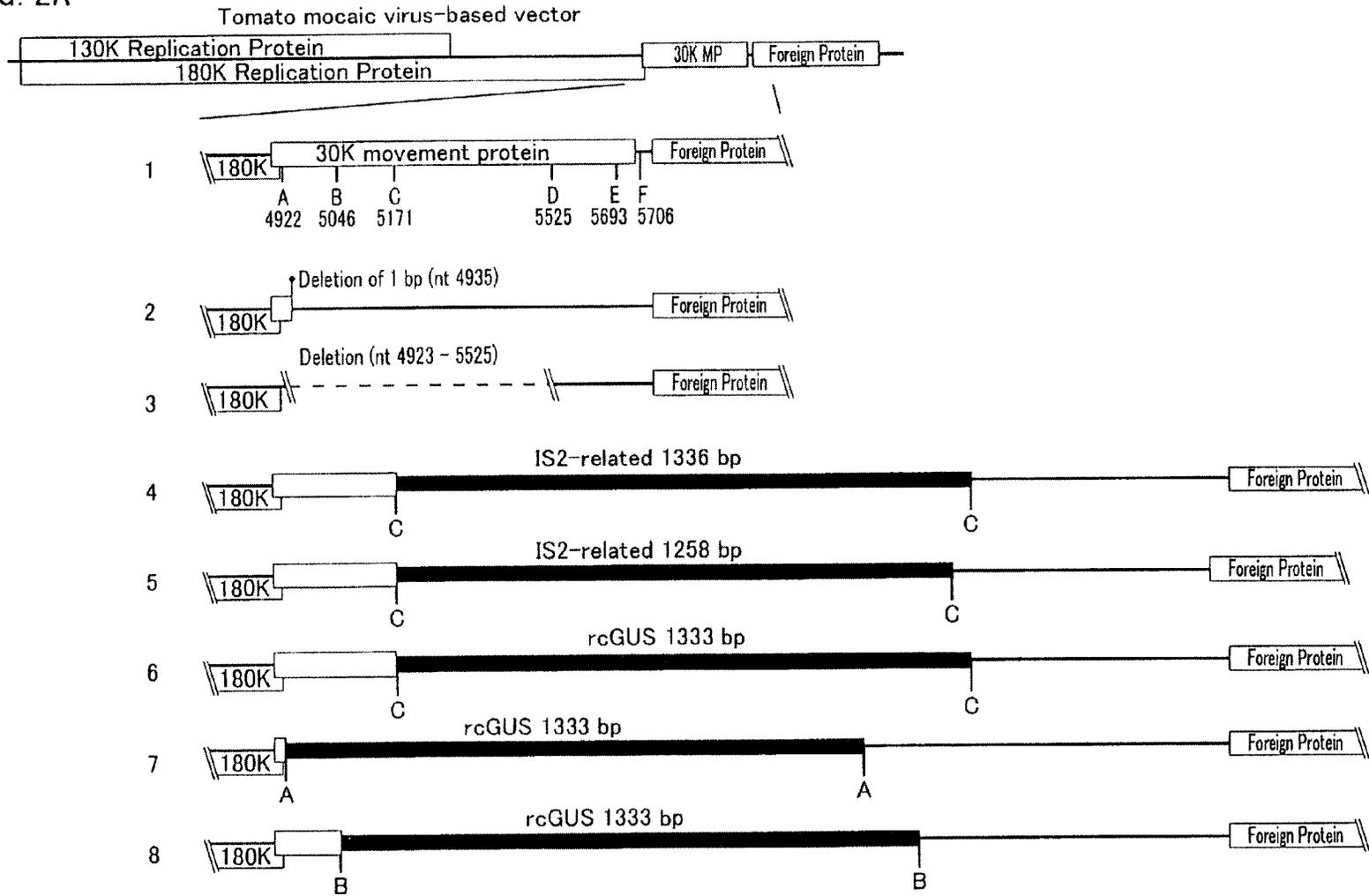
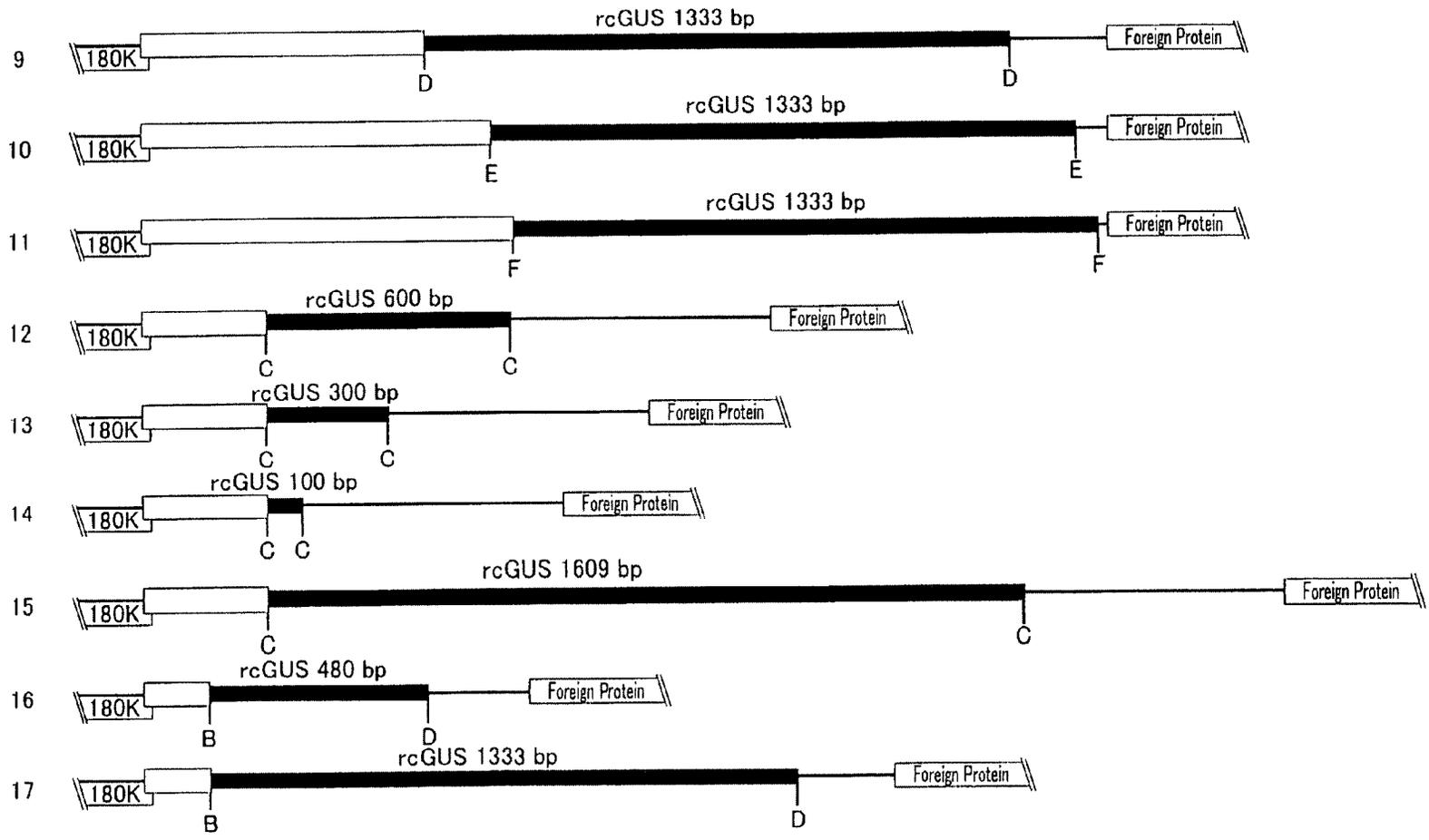


FIG. 2B



Number indicate the position in the wild-type ToMV sequence

FIG. 3

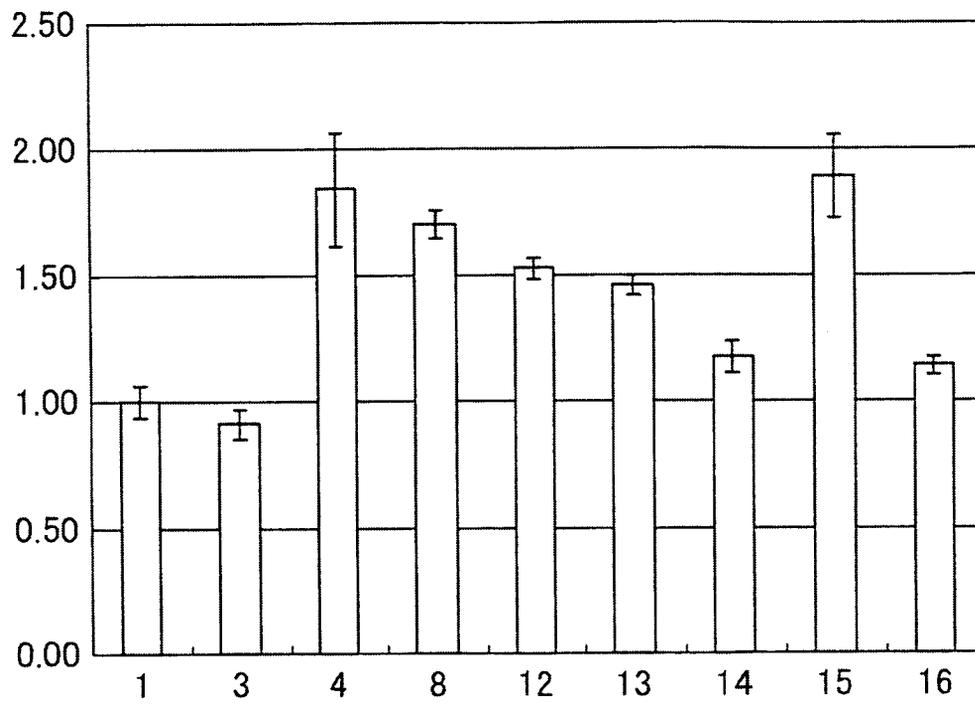
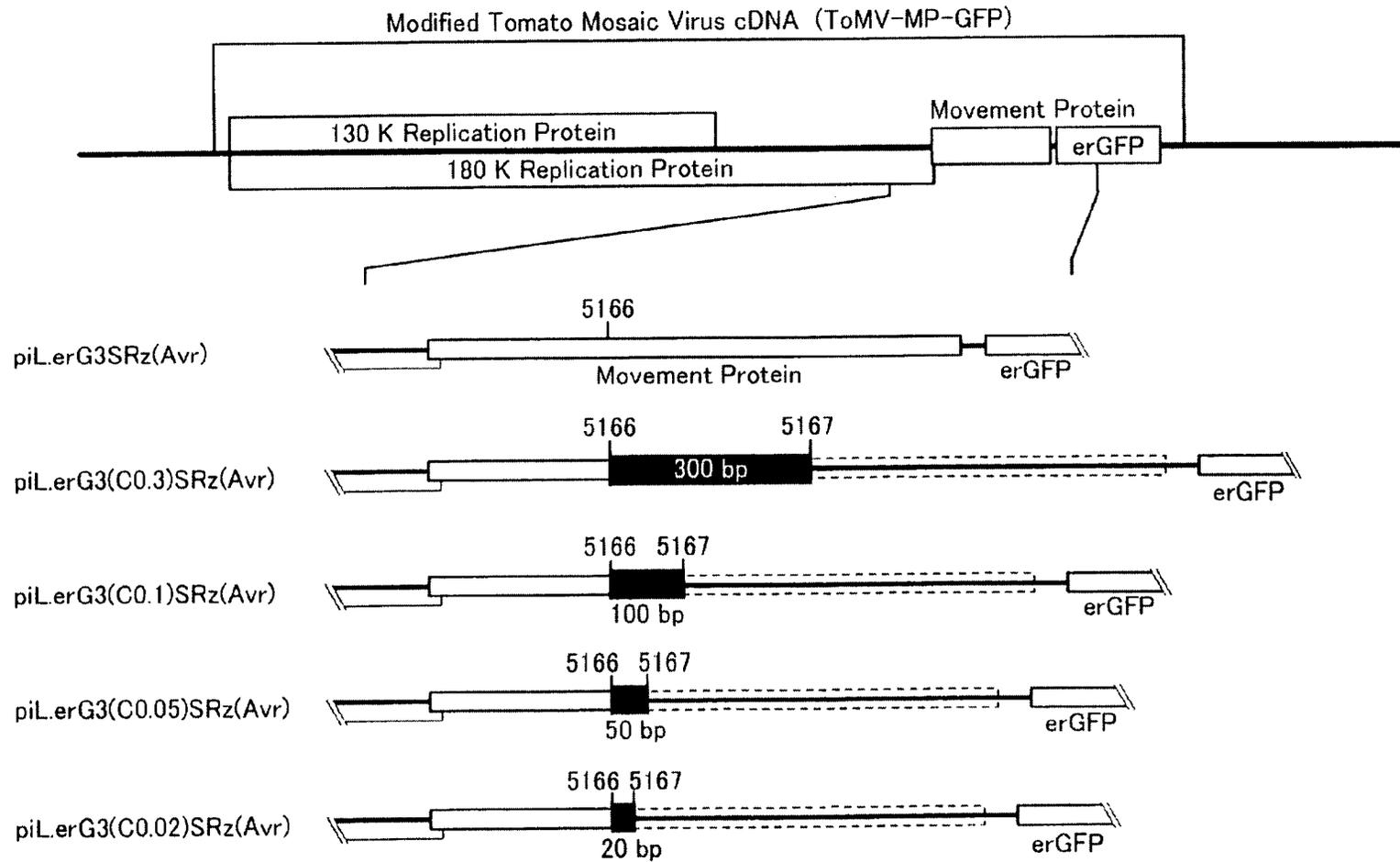


FIG. 4



## VIRUS VECTOR AND USE THEREOF

## TECHNICAL FIELD

The present invention relates to a technique for producing a protein from a polynucleotide containing a viral base sequence. More specifically, the present invention relates to (i) a polynucleotide containing a modified viral base sequence, (ii) a vector containing the polynucleotide, (iii) a plant or a transformant into which the vector is introduced, and (iv) a protein producing method and a protein producing kit, each of which utilizes the polynucleotide, the vector, the plant, or the transformant.

## BACKGROUND ART

Examples of a method for producing a useful protein in a plant includes a method of using a transformed plant in which a foreign gene is introduced into a cell, and a method of infecting a plant cell with a virus vector. The method using a virus vector is advantageous since it provides higher expression efficiency than the method using a transformed plant.

Non Patent Document 1 discloses a method for expressing a foreign gene in a plant cell by infecting the plant cell with at least two agrobacteria into which virus vectors are introduced respectively. This method eliminates the need for creating a construct for each of plural genes when combinations of various genes are tested to find a combination for encoding a useful protein. Therefore, this method is useful in analyzing functions or the like of a large number of proteins. Further, a virus vector disclosed in Non Patent Literature 1 can realize high expression speed, can be constructed at a low cost, and can eliminate steps of a conventional gene recombination process.

It is desired that a useful protein produced in a plant can be produced efficiently and in mass scale since it is used not only for food, but also for medical products. In view of this, the inventors of the present invention have constructed a system for producing a protein by using a virus vector (see Patent Literatures 1 through 3).

As another system for producing a protein by using a virus vector, Patent Literature 4 discloses a method for increasing a production amount of a useful protein by improving efficiency of producing a transcription product. Patent Literature 4 discloses a method for expressing a target protein by inserting an intron sequence into a replication sequence of a virus vector, and introducing the virus vector thus obtained into a host cell. According to this method, in which an intron region containing a lot of adenosine (A), and thymidine (T) or uracil (U) is removed from the replication sequence of the virus vector or is substituted with an intron derived from a plant cell so that (i) decomposition of the transcription product in the plant cell can be suppressed and (ii) efficiency of producing the transcription product can be improved, it is possible to increase a production amount of a target protein.

Meanwhile, each of Non Patent Literatures 2 and 3 discloses a method for improving replication efficiency of a vector which is introduced in a host cell and which contains a base sequence of potyvirus. According to the method disclosed in Non Patent Literatures 2 and 3, an intron sequence is inserted into the base sequence of potyvirus so that a transformed sequence is obtained, and a vector containing the transformed sequence is introduced into *Escherichia coli*. In the *E. coli* to which the vector is introduced, introduction of intron suppresses expression of a virus protein encoded by the base sequence of potyvirus. This controls toxic influence of

the virus on the *E. coli*, and attains better growth of the *E. coli*, thereby improving replication efficiency of the vector in the *E. coli*.

## CITATION LIST

## Patent Literature 1

Japanese Patent Application Publication, Tokukai, No. 2005-102652 A (Publication Date: Apr. 21, 2005)

## Patent Literature 2

Japanese Patent Application Publication, Tokukai, No. 2005-245228 A (Publication Date: Sep. 15, 2005)

## Patent Literature 3

Japanese Patent Application Publication, Tokukai, No. 2005-110594 A (Publication Date: Apr. 28, 2005)

## Patent Literature 4

WO2005/049839 (Publication Date: Feb. 6, 2005)

## Non Patent Literature 1

S. Marillonnet et al., PNAS, 101 (18): 6852-6857 (2004)

## Non Patent Literature 2

I. E. Johansen, PNAS. USA, 93: 12400-12405 (1996)

## Non Patent Literature 3

S. J. Yang et al., Arch Virol, 143: 2443-2451 (1998)

## SUMMARY OF INVENTION

Introduction of a vector containing a virus DNA sequence into a host cell such as *E. coli* or *agrobacterium* worsens growth of the host cell, thereby completely inhibiting the growth of the host cell, or even if the host cell can grow, the host cell grows with a poorer growth rate. This causes a reduction in yield of the vector, thereby undesirably preventing a target useful protein from being efficiently produced using the vector.

According to the conventional protein producing methods described above, it is possible to increase a production amount of a target useful protein by improving a transcriptional activity in a host cell or increasing a production amount of a transcription product. However, these methods do not take into consideration growth of a host cell into which a vector containing a virus base sequence is introduced. As such, the growth of the host cell is inhibited, and this causes a reduction in yield of the vector containing the virus base sequence. This makes it difficult to efficiently carry out genetic recombination using the vector. That is, in a case where the vector is used for protein production, a reduction in yield of the vector causes a reduction in production amount of a useful protein using the vector.

The method disclosed in Non Patent Literatures 2 and 3 allows an improvement in growth of a host cell. However, expression of a virus protein is suppressed by inserting an intron into a viral sequence. This necessitates extracting an intron sequence from each molecule of a transcription product. This causes a reduction in growth rate of virus contained

in a vector, thereby making it impossible to use the vector in efficiently producing a useful protein.

The present invention was attained in view of the above problems, and an object of the present invention is to provide a technique in which growth of a host cell, into which a vector containing a polynucleotide is introduced, is improved by using the polynucleotide containing a viral base sequence so that (i) replication efficiency of the vector in the host cell can be improved and (ii) efficiency of producing a protein using the vector can be improved.

In order to construct a virus vector, which contains a viral base sequence and does not causes deterioration in growth of a host cell, and loss of replication capability of the virus vector, the inventors of the present invention studied conditions required to construct such a virus vector. As a result of the study, the inventors of the present invention found that, in a case where a specific region of a base sequence of a tomato mosaic virus is modified, growth of a host cell, into which a vector containing the base sequence of the tomato mosaic virus is introduced, is not worsened, and that a yield of the vector in the host cell is increased accordingly. Based on this finding, the inventors of the present invention attained the present invention.

A polynucleotide of the present invention includes a viral base sequence, the viral base sequence containing: a first base sequence encoding a viral replication protein; and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, and the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition.

The virus preferably belongs to a tobomovirus. Further, the virus is preferably a tobacco mosaic virus or a tomato mosaic virus.

It is preferable that the viral replication protein is: (i) polypeptides having amino acid sequences shown in SEQ ID NO: 1 and 2, respectively, or (ii) polypeptides having amino acid sequences which are mutants of the amino acid sequences shown in SEQ ID NO: 1 and 2, respectively, or which are one of the amino acid sequences shown in SEQ ID NO: 1 and 2 and a mutant of the other, wherein mutation of the mutants is deletion, substitution, or addition of one or several amino acids therein.

It is preferable that the viral movement protein is: (i) a polypeptide having an amino acid sequence shown in SEQ ID NO: 3, or (ii) polypeptide having an amino acid sequence in which one or several amino acids are deleted, substituted, or added in the amino acid sequence shown in SEQ ID NO: 3.

It is preferable that a polynucleotide having the second base sequence is: (i) a polynucleotide having the base sequence shown in any one of SEQ ID NO: 4 through 17, (ii) a polynucleotide having a base sequence in which one or several amino acids are deleted, substituted, or added in the base sequence shown in any one of SEQ ID NO: 4 through 17, (iii) a polynucleotide which hybridizes with a polynucleotide having a base sequence that is complementary to the base sequence shown in any one of SEQ ID NO: 4 through 17 under a stringent condition, and (iv) a polynucleotide having a base sequence which has at least 80% identity with the base sequence shown in any one of SEQ ID NO: 4 through 17.

It is preferable that the base sequence with which the second base sequence is modified by the insertion, substitution, or addition has a base length of 100 or more. Further, it is preferable that the second base sequence is obtained by add-

ing the base sequence at any position from 17th base to 795th base of the base sequence shown in SEQ ID NO: 20.

A vector of the present invention contains any one of the polynucleotides.

A plant of the present invention contains any one of the polynucleotides.

A plant of the present invention contains the vector.

A transformant of the present invention contains any one of the polynucleotides.

A transformant of the present invention contains the vector.

A method of the present invention for producing a polypeptide, includes: transforming or transfecting a plant with the polynucleotide.

A method of the present invention for producing a polypeptide, includes: transforming a cell with the polynucleotide.

A kit of the present invention for producing a polypeptide, includes the polynucleotide.

A method of the present invention for producing a polypeptide, includes: transforming or transfecting a plant with the vector.

A method of the present invention for producing a polypeptide, includes: transforming a cell with the vector.

A kit of the present invention for producing a polypeptide, includes the vector.

A method of the present invention for producing a polypeptide, includes the step of: using the plant.

A method of the present invention for producing a polypeptide, includes the step of: using the transformant.

A kit of the present invention for producing a polypeptide, includes the plant.

A kit of the present invention for producing a polypeptide, includes the transformant.

For a fuller understanding of the nature and advantages of the invention; reference should be made to the ensuing detailed description taken in conjunction with the accompanying drawings.

#### BRIEF DESCRIPTION OF DRAWINGS

FIG. 1 is a view schematically illustrating a structure of a polynucleotide, according to the present invention, containing a viral base sequence.

FIG. 2A is a view schematically illustrating structures of plasmid constructs constructed in an Example of the present invention.

FIG. 2B is a view schematically illustrating structures of plasmid constructs constructed in the Example of the present invention.

FIG. 3 is a graph showing the diameter of *Escherichia coli* colonies having respective plasmid constructs in an Example of the present invention.

FIG. 4 is a view schematically illustrating structures of plasmid constructs constructed in an Example of the present invention.

#### DESCRIPTION OF EMBODIMENTS

Introduction of a vector containing a viral sequence into a host cell causes deterioration in growth of the host cell, and thereby causes a reduction in growth rate of the vector. This causes a reduction in amount of a protein that is produced using the vector and that is encoded by a foreign gene.

The inventors of the present invention aimed to construct an efficient protein producing system by constructing a vector which contains a viral sequence and which does not deteriorate growth of a host cell, and studied conditions required to construct such a vector.

[I. Polynucleotide Containing Viral Base Sequence, and Vector Containing the Polynucleotide]

The present invention provides a polynucleotide containing a viral base sequence which contains a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein.

The polynucleotide of the present invention containing the viral base sequences is a polynucleotide that is capable of functioning as a virus vector. The term "virus vector" used herein refers to a polynucleotide which contains a sequence derived from a viral genome and contains a foreign gene expressively, and preferably refers to (i) an RNA containing an RNA sequence derived from a virus, (ii) a DNA containing a cDNA sequence of an RNA derived from a virus, each of which contains a foreign gene expressively, or (ii) an RNA transcribed from this.

According to the polynucleotide of the present invention, the second base sequence is located downstream of the first base sequence. The polynucleotide of the present invention contains a part of or all of a native base sequence derived from a virus, and can be used to produce any protein in a cell.

The term "viral base sequence" used herein refers to a genome base sequence of a wild-type virus, and preferably refers to a genome RNA of an RNA virus or a cDNA obtained from the genome RNA.

The term "viral replication protein" used herein refers to a protein which is derived from a virus and which is involved in replication of a virus, and may be referred to simply as "replication protein". The protein which is involved in replication of a virus refers to a protein which replicates a virus in a cell infected with the virus. Such a protein may be a conventional replication protein, and examples of such a protein include an RNA dependent RNA polymerase (RdRp), an RNA replication enzyme, a tobacco mosaic virus 130K protein, a tobacco mosaic virus 180K protein, a tomato mosaic virus 130K protein, a tomato mosaic virus 180K protein, and the like.

A base sequence encoding the viral replication protein is preferably a native base sequence derived from a virus, but can be a base sequence which is transformed from a native base sequence derived from a virus and which encodes a protein having a replication functional activity. The term "replication functional activity" used herein refers to a functional activity of replicating a virus in a cell infected with the virus.

The term "viral movement protein" used herein refers to a protein which is derived from a virus and which is involved in intercellular movement of a virus, and may be referred to simply as "movement protein". The protein which is involved in intercellular movement of a virus refers to a protein which contributes to spread of infection of the virus by causing the virus to move from a cell infected with the virus to a neighboring cell. Such a protein may be a conventionally known movement protein, and examples of such a protein include a tobacco mosaic virus 30K protein, a tomato mosaic virus 30K protein, and the like.

A base sequence encoding the viral movement protein is preferably a native base sequence derived from a virus, but can be a base sequence which is transformed from a native base sequence derived from a virus and which encodes a protein having a movement functional activity or a base sequence which is transformed from a native base sequence derived from a virus and which encodes a protein that has lost the movement functional activity due to the transformation. The term "movement functional activity" used herein refers to a functional activity of causing a virus to move from a cell infected with the virus to a neighboring cell.

The virus is preferably a virus belonging to a tobamovirus, but is not limited to this. Examples of the virus include a tobacco mosaic virus (TMV), a tobacco mosaic virus-OM (TMV-OM), a tobacco mosaic virus-Cg (TMV-Cg), a tomato mosaic virus (ToMV), and a Sunn-hemp mosaic virus (SHMV). It should be noted that the virus is not limited to these.

The following description deals with the viral replication protein and the viral movement protein by taking a tomato mosaic virus as an example. Note that the tomato mosaic virus is a virus belonging to a tobamovirus.

Polypeptides constituting a replication protein of the tomato mosaic virus are provided as the amino acid sequences shown in SEQ ID NOs: 1 and 2, and base sequences of polynucleotide encoding the polypeptides are provided as the base sequences shown in SEQ ID NOs: 18 and 19.

In one aspect, the replication protein of the tomato mosaic virus can be (i) polypeptides respectively having the amino acid sequences shown in SEQ ID NO: 1 and 2 or (ii) polypeptides having amino acid sequences which are mutants of the amino acid sequences shown in SEQ ID NO: 1 and 2, or which are one of them and a mutant of the other one of them, each mutant polypeptide having a functional activity of replicating a virus genome.

In another aspect, the replication protein of the tomato mosaic virus can be (i) polypeptide encoded by polynucleotides respectively having the base sequences shown in SEQ ID NO: 18 and 19 or (ii) polypeptides encoded by base sequences which are mutants of the base sequences shown in SEQ ID NO: 18 and 19, or which are one of them and a mutant of the other one of them, each mutant polypeptide having a functional activity of replicating a virus genome.

That is, the replication protein of the tomato mosaic virus is constituted by two proteins, i.e., a 130K protein (referred to also as a 126K protein) having the amino acid sequence shown in SEQ ID NO: 1, and a 180K protein (referred to also as a 183K protein) having the amino acid sequence shown in SEQ ID NO: 2. The 130K protein is a direct translation product of the genome sequence of the tomato mosaic virus which is shown in SEQ ID NO: 35, and is encoded by the polynucleotide having the base sequence shown SEQ ID NO: 18. The 180K protein is a read-through translation product of the genome sequence of the tomato mosaic virus which is shown in SEQ ID NO: 35, and is encoded by a polynucleotide having the base sequence shown in SEQ ID NO: 19.

A polypeptide constituting a movement protein of the tomato mosaic virus is provided as an amino acid sequence shown in SEQ ID NO: 3, and a base sequence constituting a polynucleotide encoding the polypeptide is provided as a base sequence shown in SEQ ID NO: 20.

In one aspect, the movement protein of the tomato mosaic virus can be (i) a polypeptide having an amino acid sequence shown in SEQ ID NO: 3, (ii) a polypeptide which is a mutant of the polypeptide having an amino acid sequence shown in SEQ ID NO: 3 and which has a functional activity of causing a virus genome to move between cells, or (iii) a polypeptide which is a mutant of the polypeptide having an amino acid sequence shown in SEQ ID NO: 3 and which has lost the movement functional activity due to the mutation.

In another aspect, the movement protein of the tomato mosaic virus can be (i) a polypeptide which is encoded by a polynucleotide having a base sequence shown in SEQ ID NO: 20, (ii) a polypeptide which is encoded by a mutant of the polynucleotide having a base sequence shown in SEQ ID NO: 20 and which has a functional activity of causing a virus genome to move between cells, or (iii) a polypeptide which is encoded by a mutant of the polynucleotide having a base

sequence shown in SEQ ID NO: 20 and which has lost the movement functional activity due to the mutation.

As long as it is used in association with a protein or a polypeptide, the term “mutant” used herein refers to a polypeptide which is different in amino acid sequence, but preserves an activity of a wild-type polypeptide. That is, in this specification, a mutant of a polypeptide can be a mutant having an amino acid sequence in which one or several amino acids are deleted, substituted, or added in a specific amino acid sequence.

It is known in the art that several amino acids in an amino acid sequence of a polypeptide can be easily modified without causing a significant influence on a structure or a function of the polypeptide. Further, it is also known that mutation occurs not only in an artificially modified protein, but also in a naturally existing protein without causing a significant change in structure and function of the protein. A person skilled in the art can easily modify one or several amino acids in an amino acid sequence of a polypeptide by utilizing a known art.

The above description has discussed the viral replication protein and the viral movement protein by taking the tomato mosaic virus as an example. However, a person skilled in the art will readily understand that the virus is not limited to the tomato mosaic virus.

Note that the term “protein” is exchangeable with “peptide” or “polypeptide”. Further, the term “base sequence” is exchangeable with “nucleic acid sequence” or “nucleotide sequence”, and is expressed as a sequence of bases, i.e., adenine (A), guanine (G), cytosine (C), and thymine (T) in deoxyribonucleotide, or adenine (A), guanine (G), cytosine (C), and uracil (U) in ribonucleotide.

The polynucleotide of the present invention has a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence to a native sequence derived from a virus by insertion, substitution, or addition. That is, a base sequence is added by insertion, substitution, or addition in the second base sequence which encodes the viral movement protein and which is located downstream of the first base sequence and upstream of the linking site for linking with the exogenous base sequence.

According to the polynucleotide of the present invention, the second base sequence is provided as a polynucleotide shown in SEQ ID NO: 4 through 17 or as a mutant of the polynucleotide.

As long as it is used in association with gene or polynucleotide, the term “mutant” used herein refers to a polynucleotide encoding a polypeptide which is different in base sequence but which preserves an activity inherent in polypeptide encoded by a wild-type polynucleotide. That is, in this specification, a mutant of a polynucleotide refers to (i) a polynucleotide having a base sequence in which one or several bases are deleted, substituted, or added in a specific base sequence, (ii) a polynucleotide which hybridizes with a polynucleotide having a specific base sequence or a base sequence that is complementary to the specific base sequence under a stringent condition, or (iii) a polynucleotide having a base sequence which has at least 80% identity with a specific base sequence.

The hybridization can be carried out by a known method such as a method described in “Molecular Cloning: A Laboratory Manual 3rd Edition, J. Sambrook and D. W. Russell, Cold Spring Harbor Laboratory, NY (2001)” (the contents of which are hereby incorporated by reference).

The term “stringent condition for hybridization” used herein refers to such a condition that (i) incubation is carried out overnight at 42° C. in a hybridization solution (50% formamide, 5×SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5×Denhardt’s solution, 10% dextran sulfate, and 20 µg/ml of denatured and fragmented salmon sperm DNA); and then (ii) a filter is washed with 0.1×SSC at approximately 65° C.

The polynucleotide of the present invention is such that the linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed is located downstream of the second base sequence which is located downstream of the first base sequence. The exogenous base sequence is linked with the linking site and a cell is transformed using the polynucleotide containing the exogenous base sequence so that a polypeptide encoded by the exogenous base sequence can be expressed in the cell. Note that the exogenous base sequence linked with the polynucleotide of the present invention does not need to be located adjacently to the second base sequence.

The linking site, of the polynucleotide of the present invention, for linking with the exogenous base sequence does not need to exist as a cassette in the base sequence of the polynucleotide of the present invention, provided that the exogenous base sequence can be inserted into or linked with the base sequence of the polynucleotide of the present invention. The polynucleotide of the present invention makes it possible to amplify a gene having the exogenous base sequence linked with the linking site, and thereby makes it possible to produce a product of the gene.

According to the polynucleotide of the present invention, the second base sequence is a mutant of a native sequence derived from a virus, wherein the mutation adds a base sequence to the native sequence by insertion, substitution, or addition. The term “native sequence derived from a virus” used herein refers to a sequence indigenous in a wild-type virus. That is, such a native sequence can be a natural sequence which is obtained from a wild-type virus and which is not mutated.

The second base sequence of the present invention may be obtained by modifying with a base sequence in a mutant base sequence of a native base sequence derived from a virus which mutant base sequence encodes a polypeptide having the movement functional activity wherein the modification modifies with a base sequence in the native base sequence by insertion, substitution, or addition. Further, the second base sequence of the present invention may be obtained by modifying with a base sequence in a mutant base sequence of a native base sequence derived from a virus which mutant base sequence encodes a polypeptide having no movement functional activity due to the mutation wherein the modification modifies with a base sequence in the native base sequence by insertion, substitution, or addition.

In the polynucleotide of the present invention, the base sequence which is included in the second base sequence by insertion, substitution, or addition can be any base sequence having a certain base length, and therefore can have any sequence and can be derived from anything. In this specification, such a base sequence which is added to the second base sequence by insertion, substitution, or addition may be also referred to simply as “insertion sequence”. In later described Examples of the present invention, a sequence derived from *Escherichia coli* transposon IS2 and a sequence derived from reverse complement of a GUS gene were used as the insertion sequences. These sequences were successfully used as the insertion sequences in the Examples.

In one embodiment, an insertion sequence used in the polynucleotide of the present invention may have 100 bases or more, preferably has 100-1609 bases, and more preferably has 300-1609 bases. The inclusion of such an insertion sequence having not less than 100 base length in the second base sequence of the polynucleotide of the present invention by insertion, substitution, or addition causes a further improvement in growth of a host cell into which the polynucleotide is introduced (see the Example described later). This improves efficiency of replicating a vector in the cell, thereby allowing a further increase in yield of the vector.

The insertion sequence can be inserted in any position in a native sequence derived from a virus in order to obtain the second base sequence used in the polynucleotide of the present invention. The insertion sequence is preferably inserted in a region of the second base sequence which exists between a C-terminal region of the first base sequence and a start codon region of the exogenous base sequence linked with the linking site, and is more preferably inserted between a stop codon of the first base sequence and a subgenome promoter of a base sequence encoding a coat protein. However, the position where the insertion sequence is inserted is not limited to these. Further, the second base sequence can be obtained by adding the insertion sequence to the native sequence derived from a virus or can be obtained by substituting a part of the native sequence derived from a virus with the insertion sequence. Further, the second base sequence can be obtained by deleting a part of the native sequence derived from a virus and inserting the insertion sequence in a section where the part of the native sequence was deleted.

That is, the second base sequence of the present invention is a sequence obtained by mutating, as described above, a sequence encoding a protein which preserves a function of causing a viral genome to move between cells.

In one embodiment, the insertion sequence can be inserted in any position of the second base sequence used in the polynucleotide of the present invention. The position where the insertion sequence is inserted is not limited to a specific one, but the insertion sequence is preferably inserted in any position from 17th base to 795th base of the base sequence shown in SEQ ID NO: 20, and more preferably inserted in any position from 17th base to 620th base of the base sequence shown in SEQ ID NO: 20.

Further, the second base sequence used in the polynucleotide of the present invention may be obtained by adding the insertion sequence to 5' or 3' terminal of the base sequence shown in SEQ ID NO: 20 or may be obtained by substituting a part of the base sequence shown in SEQ ID NO: 20 with the insertion sequence.

The present invention also provides a vector for producing a polypeptide as desired. The vector of the present invention can be such a vector that contains a polynucleotide containing a viral base sequence and that is capable of expressing the polynucleotide in a host cell into which the vector is incorporated, the viral base sequence containing a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition.

A vector containing a viral base sequence can be easily mutated, and therefore construction of such a vector is very difficult, or impossible in some cases depending on the type of

a foreign gene to be expressed. However, the use of the polynucleotide of the present invention made it possible to construct such a vector that could not be constructed before. It can be estimated from this that a vector constructed using the polynucleotide of the present invention is a stable vector in which occurrence of mutation is suppressed.

Further, the vector of the present invention allows an improvement in growth of the host cell into which the vector is introduced, thereby improving replication efficiency in the host cell. Because of this, a useful protein, which is encoded by a foreign gene, can be efficiently produced by using a replicated vector.

The vector containing the polynucleotide of the present invention may be, for example, an expression vector (e.g. phage vector or plasmid vector), which can express the polynucleotide, such as a pBR type or a pUC type. A vector which can express the polynucleotide in a host cell into which the vector of the present invention is introduced can be appropriately selected as such a vector. Further, a vector which has a property of being incorporated into a genome of a plant cell can be a vector such as a pBI type or a pCAMBIA type, and can be a Ti plasmid vector, for example.

How to construct the polynucleotide of the present invention and the vector of the present invention is not limited particularly, and they may be constructed by a known genetic engineering method.

The vector constructed using the polynucleotide of the present invention can be suitably used in production of a protein encoded by a foreign gene. That is, transformation of a host cell by using a vector containing the polynucleotide of the present invention can efficiently replicate the vector without worsening growth of the host cell, thereby making it possible to efficiently produce, by using the replicated vector, the protein encoded by the foreign gene.

The foreign gene linked with the linking site contained in the polynucleotide of the present invention or the vector of the present invention is not limited to a specific one, and therefore can be a GFP gene, a human gamma interferon gene, an alpha interferon gene, a calmodulin gene, a myosin phosphatase inhibitor protein (CPI-17) functional domain gene (amino acid residue: 22-120), or a single chain antibody gene, for example. The use of the polynucleotide of the present invention or the vector of the present invention allows easy preparation of a vector carrying such a gene, efficient replication of such a vector and efficient production of a protein from the gene by using the replicated vector.

#### [2. Plant Containing Viral Base Sequence]

The present invention also provides a plant containing a viral base sequence. The term "plant" used herein refers to a plant cell or a plant individual, and examples of the plant include plants such as *Arabidopsis*, tobacco, or benthamiana, and plant cells such as a tobacco BY2 cell or an *Arabidopsis* mm2d cell.

The plant of the present invention contains a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition. With this, the polypeptide can be efficiently expressed.

In one embodiment, the plant of the present invention is obtained by introducing, into an organism, a polynucleotide containing a viral base sequence or a vector containing the

polynucleotide, the viral base sequence containing a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition.

In one aspect, the plant of the present invention may be obtained by transforming a plant or a plant cell using the polynucleotide of the present invention or a vector containing the polynucleotide of the present invention. The plant of the present invention can be obtained, for example, by introducing the polynucleotide of the present invention into a plant cell by a method such as electroporation.

In another aspect, the plant of the present invention can be obtained by transfecting a plant or a plant cell with the polynucleotide of the present invention or a vector containing the polynucleotide of the present invention. The plant of the present invention can be obtained, for example, by infecting a plant or a plant cell with the polynucleotide of the present invention. Further, the plant of the present invention can also be obtained by transfecting a plant cell with a plasmid into which cDNA obtained by adding a promoter to the polynucleotide of the present invention has been introduced and transcribing the cDNA in the cell. Further, the plant of the present invention can also be obtained by transfecting a plant cell with cDNA of the polynucleotide of the present invention and transcribing the cDNA in the cell.

Further, the plant of the present invention can also be obtained by infecting a plant cell with *agrobacterium* into which a plasmid vector containing the polynucleotide of the present invention is introduced, for example. Further, the plant of the present invention can also be obtained by agroinfiltration utilizing *agrobacterium*. Specifically, the polynucleotide of the present invention is locally introduced into a plant body by infiltrating a culture solution, in which *agrobacterium* containing the polynucleotide of the present invention is incubated, into intercellular space of the plant body.

That is, the plant of the present invention may be a transformed plant which has been transformed using the polynucleotide of the present invention or the vector of the present invention, or can be an infected plant which is infected with the polynucleotide of the present invention or the vector of the present invention. In a case where the plant of the present invention is a transformed plant, it can be a transient transformant in which the polynucleotide of the present invention which is introduced into a plant does not integrate with the genome of the plant and is transiently expressed, or can be a stable transformant in which the polynucleotide of the present invention which is introduced in a plant integrates with the genome of the plant and is stably and continuously expressed. Further, in the transformed plant, polynucleotide of the present invention which is introduced into the plant may be constantly expressed or may be inducibly expressed using steroid hormone or the like. In a case where the plant of the present invention is an infected plant, the plant may be entirely infected with the polynucleotide of the present invention or may be locally infected with the polynucleotide of the present invention.

Since the plant of the present invention contains the polynucleotide of the present invention, the use of the plant of the present invention allows efficient production of a protein

encoded by a foreign gene which is incorporated in the polynucleotide of the present invention or the vector of the present invention.

### [3. Transformant Containing Viral Base Sequence]

The present invention also provides a transformant containing a viral base sequence. The term "transformant" includes not only cell, tissue, and organ, but also individual organism, but the transformant is preferably a cell (especially prokaryotic cell, fungus, or the like). A transformant of the present invention can be *Escherichia coli*, *agrobacterium*, or yeast, for example.

The transformant of the present invention contains a polynucleotide containing a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition. As such, the transformant of the present invention can be used in efficient expression of the polypeptide.

In one embodiment, the transformant of the present invention is obtained by introducing, into an organism, a polynucleotide containing a viral base sequence or a vector containing the polynucleotide, the viral base sequence containing a first base sequence encoding a viral replication protein and a second base sequence encoding a viral movement protein, the second base sequence being located downstream of the first base sequence and having a linking site for linking with an exogenous base sequence encoding a polypeptide to be expressed, the linking site being located downstream of the second base sequence, the second base sequence being obtained by modifying with a base sequence in a native sequence derived from a virus by insertion, substitution, or addition.

In one aspect, the transformant of the present invention can be obtained by transforming an organism using the polynucleotide of the present invention or the vector containing the polynucleotide of the present invention. For example, the transformant of the present invention can be obtained by introducing a plasmid, into which the polynucleotide of the present invention is incorporated, into *Escherichia coli* by a method such as a calcium chloride method.

The use of the transformant of the present invention allows an increase in yield of the vector of the present invention which is introduced into a host cell, for example. That is, the use of the transformant of the present invention makes it possible to easily and efficiently produce a vector for producing a protein encoded by a foreign gene incorporated into the vector. Further, since growth of the transformant can be improved, the use of the transformant allows efficient production of a protein encoded by a foreign gene incorporated into the vector of the present invention.

### [4. Method and Kit for Producing Polypeptide in a Cell as Desired]

The present invention also provides (i) a method for producing a polypeptide using the polynucleotide, the vector, the plant, or the transformant, and (ii) a kit for producing a polypeptide, the kit including the polynucleotide, the vector, the plant, or the transformant.

A method of the present invention for producing a polypeptide uses the polynucleotide, the vector, the plant, or the transformant. A kit of the present invention for producing a polypeptide includes the polynucleotide, the vector, the plant,

or the transformant. Note that the term “kit” used herein means that at least one of the components is contained in another material (e.g. container).

The present invention provides a method and a kit for efficiently producing any polypeptide. The use of the method of the present invention for producing any polypeptide in a cell does not cause deterioration in growth of the cell even if a viral base sequence is introduced into the cell, thereby allowing efficient production of the polypeptide.

In one embodiment, the method of the present invention for producing a polypeptide uses the polynucleotide of the present invention or the vector containing the polynucleotide, the method including the step of transforming or transfecting a living specimen with the polynucleotide of the present invention or the vector of the present invention, wherein the living specimen may or may not be a plant body or a plant cell. A polypeptide encoded by an exogenous base sequence contained in the polynucleotide of the present invention or the vector of the present invention is expressed in the organism thus transformed or transfected in the step.

In another embodiment, the method of the present invention for producing a polypeptide uses the plant of the present invention or the transformant of the present invention, the method including the step of growing or incubating the plant of the present invention or the transformant of the present invention under a condition that a polypeptide can be expressed. A polypeptide encoded by an exogenous base sequence contained in the plant or the transformant is expressed in the plant or the transformant in the step.

As described above, the use of the method of the present invention for producing a polypeptide does not cause deterioration in growth of an organism in which a predetermined polypeptide is produced, thereby making it possible to efficiently produce the polypeptide.

A method of the present invention for introducing a polypeptide or a vector into a host is not limited to a specific one, and a conventionally known method such as an *agrobacterium* method, electroporation, a calcium phosphate method, a liposome method, or a DEAE dextran method can be suitably used as such a method. Further, an organism which is transformed or transfected with the vector of the present invention is not limited to a specific one, and therefore can be a cell derived from an animal or a cell derived from a plant. Further, a microorganism such as *Bacillus subtilis*, *Escherichia coli*, fungus, or yeast can be used as the host.

A method of the present invention for introducing the polynucleotide of the present invention or the vector of the present invention into a plant body or a cell derived from a plant is not limited to a specific one, and a method such as the *agrobacterium* method, the agroinfiltration, a polyethylene glycol method, the electroporation, or a particle gun method can be suitably used as such a method.

The kit of the present invention for producing a polypeptide includes the polynucleotide of the present invention, the vector of the present invention, the plant of the present invention, or the transformant of the present invention. In a preferable embodiment, the kit of the present invention for producing a polypeptide, including the polynucleotide of the present invention or the vector of the present invention preferably further includes a plant body or an organism to be transformed or transfected. With this arrangement, a cell is transformed or transfected using the polynucleotide of the present invention or the vector of the present invention so that a polynucleotide encoded by an exogenous base sequence contained in the polynucleotide of the present invention or the vector of the present invention can be expressed in the plant body or the organism into which the cell has been introduced.

Note that the method and the kit for producing any polypeptide in a cell is not limited to those explained above, and a person skilled in the art who read this specification can easily understand other aspects of the method and the kit for producing a polypeptide.

The following description deals with more detailed explanation of the present invention with reference to the Examples, but the present invention is not limited to these Examples, but may be altered by a skilled person within the scope of the claims and the embodiment. An embodiment based on a proper combination of technical means disclosed in different embodiments is encompassed in the technical scope of the present invention.

## EXAMPLES

### Example 1

#### Construction of Plasmid for Producing GFP

cDNAs of a tomato mosaic virus were synthesized by inserting various base sequences (SEQ ID NO: 21 through 34) into a base sequence (SEQ ID NO: 20) of a gene encoding a movement protein. The base sequence is located between a base sequence of a gene encoding a tomato mosaic virus replication protein and a base sequence of a gene encoding a target foreign protein. The cDNAs thus synthesized were used to construct plasmid constructs, respectively (see FIG. 1).

FIGS. 2A and 2B show positions of respective insertion base sequences with which the base sequence of the gene encoding the movement protein was modified by insertion, substitution, or addition. In FIGS. 2A and 2B, A through F indicate respective positions where the respective insertion base sequences were inserted, and numbers below the alphabets indicate respective positions of the respective insertion base sequences in a native sequence of the tomato mosaic virus. Note that a construct indicated by No. 16 (piLrcG11erG3SRz) was obtained by substituting a base sequence between B and D with the base sequence shown in SEQ ID NO: 33, and a construct indicated by No. 17 (piLrcG12erG3SRz) was obtained by substituting a base sequence between B and D with the base sequence shown in SEQ ID NO: 34.

In this Example, an insertion sequence added into the constructs indicated by No. 4 and No. 5 (piLIS2erG3SRz and piLIS2(-SpeI)erG3SRz) in Table 1 was a base sequence derived from *Escherichia coli* transposon IS2. Meanwhile, an insertion sequence added into each of the other constructs was a sequence derived from reverse complement of a GUS gene. Further, in this Example, a gene encoding a GFP protein was used as a gene encoding a target foreign protein.

The plasmid constructs constructed as above were used to transform *Escherichia coli*. One of colonies obtained from the transformed *E. coli* was inoculated into a 3 ml LB culture medium containing antibiotics for selection, and then was incubated at 37° C. for 20 hours with shaking. A plasmid was purified from a 1.5 ml incubation solution by an alkali SDS method. The plasmid thus purified was quantified using a DNA assay kit (Quant-it dsDNA Assay Kit (invitrogen)).

Yields of the plasmid constructs obtained in a cell was compared with a yield of a plasmid construct into which no insertion sequence was inserted, and obtained relative values are shown in Tables 1 and 2.

15

TABLE 1

Plasmid Name	Inserted Position	SEQ ID NO	Number of Inserted Bases (bp)	Yield of Plasmid (relative value)	SE
1 piLerG3SRz	—	—	—	1.0	0.1
2 piLerG3(SF3)SRz	—	—	—	0.6	0.0
3 piLAMPerG3SRz	—	—	—	0.6	0.0
4 piLIS2erG3SRz	C	21	1336	17.2	0.6
5 piLIS2(-SpeI)erG3SRz	C	22	1258	16.3	0.6
6 piLrcG1erG3SRz	A	23	1333	14.1	0.3
7 piLrcG2erG3SRz	B	24	1333	13.1	0.4
8 piLrcG3erG3SRz	C	25	1338	13.4	0.4
9 piLrcG8erG3SRz	D	26	1333	12.3	0.3
10 piLrcG9erG3SRz	E	27	1333	12.5	0.4
11 piLrcG10erG3SRz	F	28	1333	14.4	0.2
14 piLrcG6erG3SRz	C	31	100	2.2	0.1
15 piLrcG7.5erG3SRz	C	32	1604	13.6	0.7
16 piLrcG11erG3SRz	B/D	33	480	9.4	0.5
17 piLrcG12erG3SRz	B/D	34	1333	14.0	0.5

TABLE 2

Plasmid Name	Inserted Position	SEQ ID NO	Number of Inserted Bases (bp)	Yield of Plasmid (relative value)	SE
1 piLerG3SRz	—	—	—	1.0	0.1
12 piLrcG4erG3SRz	C	29	600	5.0	0.1
13 piLrcG5erG3SRz	C	30	300	4.3	0.3

As shown in Tables 1 and 2, yields of the plasmid constructs (indicated by No. 4 through No. 17, respectively) into which the base sequences respectively shown in SEQ ID NO: 21 through 34 were inserted increased by 2.2 to 17.2 times compared with the plasmid construct (indicated by No. 1) containing a native base sequence into which no insertion sequence was inserted. Note that a plasmid construct in which a gene encoding the movement protein was frameshifted (plasmid construct indicated by No. 2), and a plasmid construct in which a gene encoding the movement protein was deleted (plasmid construct indicated by No. 3) did not increase in yield.

### Example 2

#### Improvement in Growth Condition of Host Microorganism Cell

The plasmid constructs constructed in the Example 1 were used to transform *Escherichia coli* JM109 (TOYOBO). The *Escherichia coli* JM109 thus transformed was placed on an LB agar medium containing 100 µg/ml carbenicillin and was incubated at 37° C. for 18 hours. Five colonies whose growth was not affected by other colonies were randomly selected from obtained colonies, and each of the five colonies was measured in major axis.

A colony having a plasmid construct containing an insertion sequence and a colony having a plasmid construct containing no insertion sequence among the plasmid constructs constructed in the Example 1 were compared in major axis. Table 1 and FIG. 3 show obtained relative values of the major axis.

16

TABLE 3

Plasmid Name	Major Axis (relative value)
1 piLerG3SRz	1.0
3 piLAMPerG3SRz	0.91
4 piLIS2erG3SRz	1.84
8 piLrcG3erG3SRz	1.70
12 piLrcG4erG3SRz	1.53
13 piLrcG5erG3SRz	1.46
14 piLrcG6erG3SRz	1.18
15 piLrcG7.5erG3SRz	1.89
16 piLrcG11erG3SRz	1.14

As shown in Table 3, the major axis of an *Escherichia coli* colony having a plasmid construct containing an insertion sequence (plasmid construct indicated by 4, 8, 12, 13, 14, 15, or 16) was 1.14 to 1.89 times larger than the major axis of an *Escherichia coli* colony having a plasmid construct containing no insertion sequence (plasmid construct indicated by 1 or 3). This demonstrates that a growth condition of *Escherichia coli* into which a plasmid construct containing a viral base sequence was introduced was improved (see FIG. 3).

### Example 3

#### Construction of Plasmid for Production of Foreign Protein

In the Example 3, plasmid constructs were constructed by using a human gamma interferon (hIFN $\gamma$ ) gene as a gene encoding a foreign protein.

The hIFN $\gamma$  gene was amplified by the PCR method by using an AatII recognition site at the 5'-terminal site and a BstEII site at the 3'-terminal side of a GFP gene of each of the plasmid constructs constructed in the Example 1 (No. 1 (piLerG3SRz), No. 3 (piLAMPerG3SRz), No. 4 (piLIS2erG3SRz), No. 6 (piLrcG1erG3SRz), and No. 8 (piLrcG3erG3SRz) (see Table 1)). The hIFN $\gamma$  gene was then accurately substituted, so that plasmid constructs (No. 1' (piLhIFN $\gamma$ SRz), No. 3' (piLAMPhIFN $\gamma$ SRz), No. 4' (piLIS2hIFN $\gamma$ SRz), No. 6' (piLrcG1hIFN $\gamma$ SRz), and No. 8' (piLrcG3hIFN $\gamma$ SRz)) were constructed.

Yields of the plasmid constructs in respective cells was quantitatively analyzed in the same manner as in the Example 1. The result demonstrated that a yield of a plasmid construct into which an insertion base sequence was inserted (No. 4', 6', or 8') was much larger than that of a plasmid construct in which no insertion base sequence was inserted (No. 1' or 3').

It was also possible to easily construct a plasmid construct, into which a cDNA of a virus genome RNA mutated as shown in No. 4 of FIG. 2A was introduced, the virus genome RNA being mutated by using, as a gene encoding a foreign protein, an alpha interferon gene, a myosin phosphatase inhibitor protein (CPI-17) functional domain gene (amino acid residue: 22-120), a single chain antibody gene, or a calmodulin gene in a similar manner to the above Example. The plasmid construct was obtained in good yield with good stability.

### Example 4

#### Construction of Binary Plasmid

Further, each of the plasmid constructs constructed as above was cleaved with SpeI and AvrII, and was linked with a SpeI recognition site of pBICER8-ToMV5'-Spe (Dohi et al, 2006, Archives of Virology, 151: 1075-1084) in order to introduce a base sequence of a virus containing the hIFN $\gamma$  gene

17

into a binary plasmid that was to be used for inducing expression of the viral sequence therein. Although a binary plasmid into which a gene fragment derived from the plasmid construct indicated by 1' or 3' was inserted could not be obtained, a binary plasmid into which a gene fragment derived from the plasmid construct indicated by 4', 6', or 3' was inserted could be easily obtained. This revealed that a plasmid construct which contains a foreign gene and whose construction is difficult can be constructed by inserting, substituting or adding an insertion sequence in a base sequence encoding a viral movement protein.

It was also possible to easily construct a binary plasmid, into which a cDNA of a virus genome RNA mutated as shown in No. 4 of FIG. 2A was introduced, the virus genome RNA being mutated by using, as a gene encoding a foreign protein, an alpha interferon gene, a CPI-17 protein functional domain gene, a single chain antibody gene, or a calmodulin gene in a similar manner to the above Example. The binary plasmid was obtained in good yield with good stability.

#### Example 5

##### Expression of Protein in Protoplast

As shown in No. 4 through No. 15 of FIGS. 2A and 2B, an insertion sequence was inserted, substituted, or added in a virus genome RNA that was synthesized in a test tube with the use of T7RNA polymerase. Thus, a mutant of the virus genome RNA was created. The virus genome RNA thus created was inoculated into a protoplast, which was prepared from a tobacco BY2 cell, by electroporation (as for an experimental method, see Watanabe et al, FEBS Letters, 219:65-69). A transformant of the protoplast thus obtained was incubated at 26° C. for 24 hours, and then was sampled.

In a protoplast which contains a virus genome RNA into which a GFP gene was introduced as a foreign gene, proliferation of the virus genome RNA was confirmed by northern blotting. In protoplasts which respectively contain virus genome RNAs shown in No. 4 through No. 15, respectively, proliferation of the virus genome RNAs was confirmed. Further, proliferation of a sub genome GFP messenger RNA was confirmed in each of protoplasts respectively containing virus genome RNAs having respective insertion sequences shown in No. 4 through No. 9 and No. 12 through No. 15, respectively. Meanwhile, accumulation of the sub genome GFP messenger RNA could not be detected in each of protoplasts respectively containing virus genome RNAs shown in No. 10 and No. 11, respectively.

In the protoplasts, expression of a GFP gene was confirmed with the use of a fluorescent microscope. Note that expression of a GFP gene was confirmed in each of the protoplasts respectively containing the virus genome RNAs shown in No. 4 through No. 9 and No. 12 through No. 15, but expression of a GFP gene was not confirmed in each of the protoplasts respectively containing the virus genome RNAs shown in No. 10 and No. 11.

It can be estimated that the reason why the sub genome GFP messenger RNA was not accumulated in each of the protoplasts respectively containing the virus genome RNAs having insertion sequences shown in No. 10 and No. 11 lies in that a viral sub genome RNA promoter region was modified due to insertion or addition of the insertion sequences. This follows that the GFP gene can be expressed also in these virus genome RNAs by further adding a native sub genome RNA promoter sequence.

Further, in a protoplast which contains a virus genome RNA into which the hIFN $\gamma$  gene was introduced as a foreign

18

gene, proliferation capability of the virus genome RNA was confirmed by northern blotting, and expression of the hIFN $\gamma$  gene was confirmed by western blotting. In protoplasts which respectively contain the virus genome RNAs having insertion sequences shown in No. 6, No. 7, and No. 12 through No. 14, proliferation of the virus genome RNAs and proliferation of a sub genome hIFN $\gamma$  messenger RNA was confirmed, and expression of the hIFN $\gamma$  gene was confirmed since a hIFN $\gamma$  protein was detected.

Similarly, proliferation of a genome RNA and a sub genome messenger RNA was confirmed in a protoplast containing a mutant of a virus genome into which a cDNA of a virus genome RNA mutated as shown in No. 4 of FIG. 2A was introduced, the virus genome RNA being mutated by using, as a gene encoding a foreign protein, an alpha interferon gene, a CPI-17 protein functional domain gene, a single chain antibody gene, or a calmodulin gene in a similar manner to the above Example.

#### Example 6

##### Expression of Protein in Tobacco BY2 Cell

A cDNA of a virus genome RNA into which a GFP gene or a hIFN $\gamma$  gene was introduced as a foreign gene (see No. 4 of FIG. 2A) was used to transform a tobacco BY2 cell (Dohi et al., Archives of Virology, 151, 1075-1084) in which a transcription factor XVE that was activated by estrogen was expressed with the use of the *agrobacterium* method. Estrogen was added to a culture medium containing the tobacco BY2 cell thus transformed, and three days later, a sample was taken (as for an experimental method, see Dohi et al., Archives of Virology, 151, 1075-1084).

In a transformed tobacco BY2 cell containing the GFP gene, proliferation of a virus genome RNA and a sub genome GFP messenger RNA was confirmed (northern blotting), and expression of the GFP gene was confirmed (fluorescence microscope observation and SDS-PAGE).

Also in a transformed tobacco BY2 cell containing the hIFN $\gamma$  gene, proliferation of a virus genome RNA and a sub genome hIFN $\gamma$  messenger RNA was confirmed (northern blotting), and accumulation of a hIFN $\gamma$  protein was confirmed (western blotting).

Further, proliferation of a virus genome RNA and a sub genome messenger RNA was confirmed, and accumulation of a protein was confirmed in a transformed tobacco BY2 cell into which a cDNA of a virus genome RNA mutated as shown in No. 4 of FIG. 2A was introduced, the virus genome RNA being mutated by using, as a gene encoding a foreign protein, a CPI-17 protein functional domain gene or a single chain antibody gene in a similar manner to the above Example.

#### Example 7

##### Study on Base Length of Insertion Sequence

cDNAs of a modified tomato mosaic virus were synthesized by inserting base sequences having base length of 300 base pairs, 100 base pairs, 50 base pairs, and 20 base pairs (SEQ ID NO: 36 through 39) at a position of 5166 bases from the 5' terminal of a base sequence (SEQ ID NO: 20) of a gene encoding a movement protein of a tomato mosaic virus. A cDNA of a modified tomato mosaic virus encoded by a plasmid vector piL.erG3SRz(Avr) was substituted with the cDNAs thus synthesized so that plasmid constructs were constructed (see FIG. 4). In FIG. 4, a plasmid construct into which no insertion sequence was inserted is indicated by

piL.erG3SRz(Avr), and plasmid constructs into which insertion sequences of 300 base length, 100 base length, 50 base length, and 20 base length were inserted are indicated by piL.erG3(C0.3)SRz(Avr), piL.erG3(C0.1)SRz(Avr), piL.erG3(C0.05)SRz(Avr), piL.erG3(C0.02)SRz(Avr), respectively.

The plasmid constructs thus constructed were used to transform *Escherichia coli* JM109 (TOYOBO). The *Escherichia coli* JM109 thus transformed was placed on an LB agar medium containing 100 µg/ml carbenicillin and was incubated at 37° C. for 26 hours. Five colonies whose growth was not affected by other colonies were randomly selected from obtained colonies, and the diameter of each of the five colonies was measured.

The diameter of a colony of *Escherichia coli* having a plasmid containing an insertion sequence was compared with the diameter of a colony of *Escherichia coli* having a plasmid containing no insertion sequence. Table 4 shows obtained relative values and standard errors (n=5). In Table 4, "\*" indicates that the t-test revealed that there is a significant difference in colony diameter between a plasmid containing an insertion sequence and piL.erG3SRz(Avr) containing no insertion sequence.

TABLE 4

Plasmid Name	Number of Inserted Bases (bp)	Colony Diameter (relative value ± S.E.)	Yield of Plasmid (relative value ± S.E.)
piL.erG3SRz(Avr)	—	1.00 ± 0.05	1.00 ± 0.06
piL.erG3(C0.3)SRz(Avr)	300	1.59 ± 0.05*	3.87 ± 0.36*
piL.erG3(C0.1)SRz(Avr)	100	1.17 ± 0.03*	1.53 ± 0.10*
piL.erG3(C0.05)SRz(Avr)	50	1.10 ± 0.02	1.27 ± 0.08
piL.erG3(C0.02)SRz(Avr)	20	1.10 ± 0.04	1.33 ± 0.15

As shown in Table 4, the colony diameter of *Escherichia coli* having piL.erG3(C0.3)SRz(Avr) into which an insertion sequence of 300 base length was inserted or piL.erG3(C0.1)SRz(Avr) into which an insertion sequence of 100 base length was inserted is significantly larger than that of *Escherichia coli* having piL.erG3SRz(Avr) into which no insertion sequence was inserted. That is, an improvement could be observed in growth of *Escherichia coli* having piL.erG3(C0.3)SRz(Avr) and piL.erG3(C0.1)SRz(Avr). Meanwhile, the colony diameter of *Escherichia coli* having piL.erG3(C0.05)SRz(Avr) into which an insertion sequence of 50 base length was inserted or piL.erG3(C0.02)SRz(Avr) into which an insertion sequence of 10 base length was inserted is larger than that of *Escherichia coli* having piL.erG3SRz(Avr) into which no insertion sequence was inserted, but the difference was not significant.

Further, the plasmid constructs were used to transform *Escherichia coli* JM109. One of colonies obtained from the transformed *E. coli* was inoculated into a 3 ml LB culture medium containing antibiotics for selection, and then was incubated at 37° C. for 24 hours with shaking. A plasmid was purified from a 1.5 ml incubation solution by an alkali SDS method. The plasmid thus purified was quantified using a DNA assay kit (Quant-it dsDNA Assay Kit (invitrogen)). A yield of each of the plasmid constructs into which an insertion sequence was inserted was compared to that of a plasmid construct into which no insertion sequence was inserted. Table 4 shows obtained relative values and standard errors (n=3).

As shown in Table 4, a yield of *Escherichia coli* having piL.erG3(C0.3)SRz(Avr) into which an insertion sequence of 300 base length was inserted or piL.erG3(C0.1)SRz(Avr) into which an insertion sequence of 100 base length was inserted was significantly larger than that of *Escherichia coli* having piL.erG3SRz(Avr) into which no insertion sequence was inserted. This means that these plasmids showed good stability. Meanwhile, a yield of *Escherichia coli* having piL.erG3(C0.05)SRz(Avr) into which an insertion sequence of 50 base length was inserted or piL.erG3(C0.02)SRz(Avr) into which an insertion sequence of 10 base length was inserted was larger than that of *Escherichia coli* having piL.erG3SRz(Avr) into which no insertion sequence was inserted, but the difference was not significant.

The use of the present invention allows an improvement in growth of a host cell into which a vector containing a polynucleotide containing a viral base sequence is introduced, thereby improving efficiency of replicating the vector in the cell. As a result, it becomes possible to efficiently produce a useful protein with the use of a vector that is efficiently replicated.

The embodiments and concrete examples of implementation discussed in the foregoing detailed explanation serve solely to illustrate the technical details of the present invention, which should not be narrowly interpreted within the limits of such embodiments and concrete examples, but rather may be applied in many variations within the spirit of the present invention, provided such variations do not exceed the scope of the patent claims set forth below.

## INDUSTRIAL APPLICABILITY

The use of the present invention makes it possible to efficiently produce any protein, and a protein produced with the use of the present invention can be effectively applied to various fields such as plant biotechnology industry, pharmaceutical industry, and food industry.

## SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 39

<210> SEQ ID NO 1

<211> LENGTH: 1116

<212> TYPE: PRT

<213> ORGANISM: Tomato mosaic virus

<400> SEQUENCE: 1

Met Ala Tyr Thr Gln Thr Ala Thr Ser Ser Ala Leu Leu Glu Thr Val  
1 5 10 15

Arg Gly Asn Asn Thr Leu Val Asn Asp Leu Ala Lys Arg Arg Leu Tyr

-continued

20					25					30					
Asp	Thr	Ala	Val	Asp	Glu	Phe	Asn	Ala	Arg	Asp	Arg	Arg	Pro	Lys	Val
		35					40					45			
Asn	Phe	Ser	Lys	Val	Val	Ser	Glu	Glu	Gln	Thr	Leu	Ile	Ala	Thr	Lys
	50					55					60				
Ala	Tyr	Pro	Glu	Phe	Gln	Ile	Thr	Phe	Tyr	Asn	Thr	Gln	Asn	Ala	Val
	65					70					75				80
His	Ser	Leu	Ala	Gly	Gly	Leu	Arg	Ser	Leu	Glu	Leu	Glu	Tyr	Leu	Met
				85					90					95	
Met	Gln	Ile	Pro	Tyr	Gly	Ser	Leu	Thr	Tyr	Asp	Ile	Gly	Gly	Asn	Phe
			100					105					110		
Ala	Ser	His	Leu	Phe	Lys	Gly	Arg	Ala	Tyr	Val	His	Cys	Cys	Met	Pro
			115				120					125			
Asn	Leu	Asp	Val	Arg	Asp	Ile	Met	Arg	His	Glu	Gly	Gln	Lys	Asp	Ser
	130					135					140				
Ile	Glu	Leu	Tyr	Leu	Ser	Arg	Leu	Glu	Arg	Gly	Asn	Lys	His	Val	Pro
	145					150					155				160
Asn	Phe	Gln	Lys	Glu	Ala	Phe	Asp	Arg	Tyr	Ala	Glu	Met	Pro	Asn	Glu
				165					170					175	
Val	Val	Cys	His	Asp	Thr	Phe	Gln	Thr	Cys	Arg	His	Ser	Gln	Glu	Cys
			180					185					190		
Tyr	Thr	Gly	Arg	Val	Tyr	Ala	Ile	Ala	Leu	His	Ser	Ile	Tyr	Asp	Ile
		195				200						205			
Pro	Ala	Asp	Glu	Phe	Gly	Ala	Ala	Leu	Leu	Arg	Lys	Asn	Val	His	Val
	210					215					220				
Cys	Tyr	Ala	Ala	Phe	His	Phe	Ser	Glu	Asn	Leu	Leu	Leu	Glu	Asp	Ser
	225					230					235				240
His	Val	Asn	Leu	Asp	Glu	Ile	Asn	Ala	Cys	Phe	Gln	Arg	Asp	Gly	Asp
				245					250					255	
Arg	Leu	Thr	Phe	Ser	Phe	Ala	Ser	Glu	Ser	Thr	Leu	Asn	Tyr	Ser	His
			260					265					270		
Ser	Tyr	Ser	Asn	Ile	Leu	Lys	Tyr	Val	Cys	Lys	Thr	Tyr	Phe	Pro	Ala
		275					280					285			
Ser	Asn	Arg	Glu	Val	Tyr	Met	Lys	Glu	Phe	Leu	Val	Thr	Arg	Val	Asn
	290					295					300				
Thr	Trp	Phe	Cys	Lys	Phe	Ser	Arg	Ile	Asp	Thr	Phe	Leu	Leu	Tyr	Lys
	305					310					315				320
Gly	Val	Ala	His	Lys	Gly	Val	Asp	Ser	Glu	Gln	Phe	Tyr	Lys	Ala	Met
				325					330					335	
Glu	Asp	Ala	Trp	His	Tyr	Lys	Lys	Thr	Leu	Ala	Met	Cys	Asn	Ser	Glu
		340						345					350		
Arg	Ile	Leu	Leu	Glu	Asp	Ser	Ser	Ser	Val	Asn	Tyr	Trp	Phe	Pro	Lys
		355					360					365			
Met	Arg	Asp	Met	Val	Ile	Val	Pro	Leu	Phe	Asp	Ile	Ser	Leu	Glu	Thr
	370					375					380				
Ser	Lys	Arg	Thr	Arg	Lys	Glu	Val	Leu	Val	Ser	Lys	Asp	Phe	Val	Tyr
	385					390					395				400
Thr	Val	Leu	Asn	His	Ile	Arg	Thr	Tyr	Gln	Ala	Lys	Ala	Leu	Thr	Tyr
				405					410					415	
Ser	Asn	Val	Leu	Ser	Phe	Val	Glu	Ser	Ile	Arg	Ser	Arg	Val	Ile	Ile
		420						425					430		
Asn	Gly	Val	Thr	Ala	Arg	Ser	Glu	Trp	Asp	Val	Asp	Lys	Ser	Leu	Leu
		435					440					445			

-continued

---

Gln	Ser	Leu	Ser	Met	Thr	Phe	Phe	Leu	His	Thr	Lys	Leu	Ala	Val	Leu
450						455					460				
Lys	Asp	Asp	Leu	Leu	Ile	Ser	Lys	Phe	Ala	Leu	Gly	Pro	Lys	Thr	Val
465					470					475					480
Ser	Gln	His	Val	Trp	Asp	Glu	Ile	Ser	Leu	Ala	Phe	Gly	Asn	Ala	Phe
				485					490					495	
Pro	Ser	Ile	Lys	Glu	Arg	Leu	Ile	Asn	Arg	Lys	Leu	Ile	Lys	Ile	Thr
			500					505					510		
Glu	Asn	Ala	Leu	Glu	Ile	Arg	Val	Pro	Asp	Leu	Tyr	Val	Thr	Phe	His
		515					520					525			
Asp	Arg	Leu	Val	Ser	Glu	Tyr	Lys	Met	Ser	Val	Asp	Met	Pro	Val	Leu
		530				535					540				
Asp	Ile	Arg	Lys	Lys	Met	Glu	Glu	Thr	Glu	Glu	Met	Tyr	Asn	Ala	Leu
545					550					555					560
Ser	Glu	Leu	Ser	Val	Leu	Lys	Asn	Ser	Asp	Lys	Phe	Asp	Val	Asp	Val
				565					570					575	
Phe	Ser	Gln	Met	Cys	Gln	Ser	Leu	Glu	Val	Asp	Pro	Met	Thr	Ala	Ala
			580					585					590		
Lys	Val	Ile	Val	Ala	Val	Met	Ser	Asn	Glu	Ser	Gly	Leu	Thr	Leu	Thr
		595						600				605			
Phe	Glu	Gln	Pro	Thr	Glu	Ala	Asn	Val	Ala	Leu	Ala	Leu	Gln	Asp	Ser
		610				615					620				
Glu	Lys	Ala	Ser	Asp	Gly	Ala	Leu	Val	Val	Thr	Ser	Arg	Asp	Val	Glu
625					630					635					640
Glu	Pro	Ser	Ile	Lys	Gly	Ser	Met	Ala	Arg	Gly	Glu	Leu	Gln	Leu	Ala
				645					650					655	
Gly	Leu	Ser	Gly	Asp	Val	Pro	Glu	Ser	Ser	Tyr	Thr	Arg	Ser	Glu	Glu
			660					665					670		
Ile	Glu	Ser	Leu	Glu	Gln	Phe	His	Met	Ala	Thr	Ala	Ser	Ser	Leu	Ile
		675						680					685		
His	Lys	Gln	Met	Cys	Ser	Ile	Val	Tyr	Thr	Gly	Pro	Leu	Lys	Val	Gln
		690				695					700				
Gln	Met	Lys	Asn	Phe	Ile	Asp	Ser	Leu	Val	Ala	Ser	Leu	Ser	Ala	Ala
705					710					715					720
Val	Ser	Asn	Leu	Val	Lys	Ile	Leu	Lys	Asp	Thr	Ala	Ala	Ile	Asp	Leu
				725					730					735	
Glu	Thr	Arg	Gln	Lys	Phe	Gly	Val	Leu	Asp	Val	Ala	Ser	Lys	Arg	Trp
			740					745					750		
Leu	Val	Lys	Pro	Ser	Ala	Lys	Asn	His	Ala	Trp	Gly	Val	Val	Glu	Thr
			755					760					765		
His	Ala	Arg	Lys	Tyr	His	Val	Ala	Leu	Leu	Glu	His	Asp	Glu	Phe	Gly
		770				775						780			
Ile	Ile	Thr	Cys	Asp	Asn	Trp	Arg	Arg	Val	Ala	Val	Ser	Ser	Glu	Ser
785					790					795					800
Val	Val	Tyr	Ser	Asp	Met	Ala	Lys	Leu	Arg	Thr	Leu	Arg	Arg	Leu	Leu
				805					810					815	
Lys	Asp	Gly	Glu	Pro	His	Val	Ser	Ser	Ala	Lys	Val	Val	Leu	Val	Asp
			820					825					830		
Gly	Val	Pro	Gly	Cys	Gly	Lys	Thr	Lys	Glu	Ile	Leu	Ser	Arg	Val	Asn
			835					840					845		
Phe	Glu	Glu	Asp	Leu	Ile	Leu	Val	Pro	Gly	Arg	Gln	Ala	Ala	Glu	Met
			850					855				860			
Ile	Arg	Arg	Arg	Ala	Asn	Ala	Ser	Gly	Ile	Ile	Val	Ala	Thr	Lys	Asp
865					870						875				880

-continued

---

Asn Val Arg Thr Val Asp Ser Phe Leu Met Asn Tyr Gly Lys Gly Ala  
                   885                                  890                                  895  
 Arg Cys Gln Phe Lys Arg Leu Phe Ile Asp Glu Gly Leu Met Leu His  
                   900                                  905                                  910  
 Thr Gly Cys Val Asn Phe Leu Val Glu Met Ser Leu Cys Asp Ile Ala  
                   915                                  920                                  925  
 Tyr Val Tyr Gly Asp Thr Gln Gln Ile Pro Tyr Ile Asn Arg Val Thr  
                   930                                  935                                  940  
 Gly Phe Pro Tyr Pro Ala His Phe Ala Lys Leu Glu Val Asp Glu Val  
                   945                                  950                                  955                                  960  
 Glu Thr Arg Arg Thr Thr Leu Arg Cys Pro Ala Asp Val Thr His Phe  
                   965                                  970                                  975  
 Leu Asn Gln Arg Tyr Glu Gly His Val Met Cys Thr Ser Ser Glu Lys  
                   980                                  985                                  990  
 Lys Ser Val Ser Gln Glu Met Val Ser Gly Ala Ala Ser Ile Asn Pro  
                   995                                  1000                                  1005  
 Val Ser Lys Pro Leu Lys Gly Lys Ile Leu Thr Phe Thr Gln Ser  
                   1010                                  1015                                  1020  
 Asp Lys Glu Ala Leu Leu Ser Arg Gly Tyr Ala Asp Val His Thr  
                   1025                                  1030                                  1035  
 Val His Glu Val Gln Gly Glu Thr Tyr Ala Asp Val Ser Leu Val  
                   1040                                  1045                                  1050  
 Arg Leu Thr Pro Thr Pro Val Ser Ile Ile Ala Arg Asp Ser Pro  
                   1055                                  1060                                  1065  
 His Val Leu Val Ser Leu Ser Arg His Thr Lys Ser Leu Lys Tyr  
                   1070                                  1075                                  1080  
 Tyr Thr Val Val Met Asp Pro Leu Val Ser Ile Ile Arg Asp Leu  
                   1085                                  1090                                  1095  
 Glu Arg Val Ser Ser Tyr Leu Leu Asp Met Tyr Lys Val Asp Ala  
                   1100                                  1105                                  1110  
 Gly Thr Gln  
                   1115

&lt;210&gt; SEQ ID NO 2

&lt;211&gt; LENGTH: 1616

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Tomato mosaic virus

&lt;400&gt; SEQUENCE: 2

Met Ala Tyr Thr Gln Thr Ala Thr Ser Ser Ala Leu Leu Glu Thr Val  
 1                  5                                  10                                  15  
 Arg Gly Asn Asn Thr Leu Val Asn Asp Leu Ala Lys Arg Arg Leu Tyr  
                   20                                  25                                  30  
 Asp Thr Ala Val Asp Glu Phe Asn Ala Arg Asp Arg Arg Pro Lys Val  
                   35                                  40                                  45  
 Asn Phe Ser Lys Val Val Ser Glu Glu Gln Thr Leu Ile Ala Thr Lys  
                   50                                  55                                  60  
 Ala Tyr Pro Glu Phe Gln Ile Thr Phe Tyr Asn Thr Gln Asn Ala Val  
                   65                                  70                                  75                                  80  
 His Ser Leu Ala Gly Gly Leu Arg Ser Leu Glu Leu Glu Tyr Leu Met  
                   85                                  90                                  95  
 Met Gln Ile Pro Tyr Gly Ser Leu Thr Tyr Asp Ile Gly Gly Asn Phe  
                   100                                  105                                  110  
 Ala Ser His Leu Phe Lys Gly Arg Ala Tyr Val His Cys Cys Met Pro  
                   115                                  120                                  125

-continued

---

Asn Leu Asp Val Arg Asp Ile Met Arg His Glu Gly Gln Lys Asp Ser  
 130 135 140  
 Ile Glu Leu Tyr Leu Ser Arg Leu Glu Arg Gly Asn Lys His Val Pro  
 145 150 155 160  
 Asn Phe Gln Lys Glu Ala Phe Asp Arg Tyr Ala Glu Met Pro Asn Glu  
 165 170 175  
 Val Val Cys His Asp Thr Phe Gln Thr Cys Arg His Ser Gln Glu Cys  
 180 185 190  
 Tyr Thr Gly Arg Val Tyr Ala Ile Ala Leu His Ser Ile Tyr Asp Ile  
 195 200 205  
 Pro Ala Asp Glu Phe Gly Ala Ala Leu Leu Arg Lys Asn Val His Val  
 210 215 220  
 Cys Tyr Ala Ala Phe His Phe Ser Glu Asn Leu Leu Leu Glu Asp Ser  
 225 230 235 240  
 His Val Asn Leu Asp Glu Ile Asn Ala Cys Phe Gln Arg Asp Gly Asp  
 245 250 255  
 Arg Leu Thr Phe Ser Phe Ala Ser Glu Ser Thr Leu Asn Tyr Ser His  
 260 265 270  
 Ser Tyr Ser Asn Ile Leu Lys Tyr Val Cys Lys Thr Tyr Phe Pro Ala  
 275 280 285  
 Ser Asn Arg Glu Val Tyr Met Lys Glu Phe Leu Val Thr Arg Val Asn  
 290 295 300  
 Thr Trp Phe Cys Lys Phe Ser Arg Ile Asp Thr Phe Leu Leu Tyr Lys  
 305 310 315 320  
 Gly Val Ala His Lys Gly Val Asp Ser Glu Gln Phe Tyr Lys Ala Met  
 325 330 335  
 Glu Asp Ala Trp His Tyr Lys Lys Thr Leu Ala Met Cys Asn Ser Glu  
 340 345 350  
 Arg Ile Leu Leu Glu Asp Ser Ser Ser Val Asn Tyr Trp Phe Pro Lys  
 355 360 365  
 Met Arg Asp Met Val Ile Val Pro Leu Phe Asp Ile Ser Leu Glu Thr  
 370 375 380  
 Ser Lys Arg Thr Arg Lys Glu Val Leu Val Ser Lys Asp Phe Val Tyr  
 385 390 395 400  
 Thr Val Leu Asn His Ile Arg Thr Tyr Gln Ala Lys Ala Leu Thr Tyr  
 405 410 415  
 Ser Asn Val Leu Ser Phe Val Glu Ser Ile Arg Ser Arg Val Ile Ile  
 420 425 430  
 Asn Gly Val Thr Ala Arg Ser Glu Trp Asp Val Asp Lys Ser Leu Leu  
 435 440 445  
 Gln Ser Leu Ser Met Thr Phe Phe Leu His Thr Lys Leu Ala Val Leu  
 450 455 460  
 Lys Asp Asp Leu Leu Ile Ser Lys Phe Ala Leu Gly Pro Lys Thr Val  
 465 470 475 480  
 Ser Gln His Val Trp Asp Glu Ile Ser Leu Ala Phe Gly Asn Ala Phe  
 485 490 495  
 Pro Ser Ile Lys Glu Arg Leu Ile Asn Arg Lys Leu Ile Lys Ile Thr  
 500 505 510  
 Glu Asn Ala Leu Glu Ile Arg Val Pro Asp Leu Tyr Val Thr Phe His  
 515 520 525  
 Asp Arg Leu Val Ser Glu Tyr Lys Met Ser Val Asp Met Pro Val Leu  
 530 535 540  
 Asp Ile Arg Lys Lys Met Glu Glu Thr Glu Glu Met Tyr Asn Ala Leu

-continued

---

545		550		555		560									
Ser	Glu	Leu	Ser	Val	Leu	Lys	Asn	Ser	Asp	Lys	Phe	Asp	Val	Asp	Val
				565					570					575	
Phe	Ser	Gln	Met	Cys	Gln	Ser	Leu	Glu	Val	Asp	Pro	Met	Thr	Ala	Ala
			580					585					590		
Lys	Val	Ile	Val	Ala	Val	Met	Ser	Asn	Glu	Ser	Gly	Leu	Thr	Leu	Thr
		595					600					605			
Phe	Glu	Gln	Pro	Thr	Glu	Ala	Asn	Val	Ala	Leu	Ala	Leu	Gln	Asp	Ser
	610					615					620				
Glu	Lys	Ala	Ser	Asp	Gly	Ala	Leu	Val	Val	Thr	Ser	Arg	Asp	Val	Glu
625					630					635					640
Glu	Pro	Ser	Ile	Lys	Gly	Ser	Met	Ala	Arg	Gly	Glu	Leu	Gln	Leu	Ala
				645					650						655
Gly	Leu	Ser	Gly	Asp	Val	Pro	Glu	Ser	Ser	Tyr	Thr	Arg	Ser	Glu	Glu
		660						665					670		
Ile	Glu	Ser	Leu	Glu	Gln	Phe	His	Met	Ala	Thr	Ala	Ser	Ser	Leu	Ile
		675					680						685		
His	Lys	Gln	Met	Cys	Ser	Ile	Val	Tyr	Thr	Gly	Pro	Leu	Lys	Val	Gln
	690					695					700				
Gln	Met	Lys	Asn	Phe	Ile	Asp	Ser	Leu	Val	Ala	Ser	Leu	Ser	Ala	Ala
705				710						715					720
Val	Ser	Asn	Leu	Val	Lys	Ile	Leu	Lys	Asp	Thr	Ala	Ala	Ile	Asp	Leu
			725						730					735	
Glu	Thr	Arg	Gln	Lys	Phe	Gly	Val	Leu	Asp	Val	Ala	Ser	Lys	Arg	Trp
			740				745						750		
Leu	Val	Lys	Pro	Ser	Ala	Lys	Asn	His	Ala	Trp	Gly	Val	Val	Glu	Thr
		755					760					765			
His	Ala	Arg	Lys	Tyr	His	Val	Ala	Leu	Leu	Glu	His	Asp	Glu	Phe	Gly
	770					775					780				
Ile	Ile	Thr	Cys	Asp	Asn	Trp	Arg	Arg	Val	Ala	Val	Ser	Ser	Glu	Ser
785				790						795					800
Val	Val	Tyr	Ser	Asp	Met	Ala	Lys	Leu	Arg	Thr	Leu	Arg	Arg	Leu	Leu
			805						810					815	
Lys	Asp	Gly	Glu	Pro	His	Val	Ser	Ser	Ala	Lys	Val	Val	Leu	Val	Asp
			820					825					830		
Gly	Val	Pro	Gly	Cys	Gly	Lys	Thr	Lys	Glu	Ile	Leu	Ser	Arg	Val	Asn
		835					840					845			
Phe	Glu	Glu	Asp	Leu	Ile	Leu	Val	Pro	Gly	Arg	Gln	Ala	Ala	Glu	Met
	850					855					860				
Ile	Arg	Arg	Arg	Ala	Asn	Ala	Ser	Gly	Ile	Ile	Val	Ala	Thr	Lys	Asp
865					870					875					880
Asn	Val	Arg	Thr	Val	Asp	Ser	Phe	Leu	Met	Asn	Tyr	Gly	Lys	Gly	Ala
			885						890					895	
Arg	Cys	Gln	Phe	Lys	Arg	Leu	Phe	Ile	Asp	Glu	Gly	Leu	Met	Leu	His
			900					905					910		
Thr	Gly	Cys	Val	Asn	Phe	Leu	Val	Glu	Met	Ser	Leu	Cys	Asp	Ile	Ala
		915					920						925		
Tyr	Val	Tyr	Gly	Asp	Thr	Gln	Gln	Ile	Pro	Tyr	Ile	Asn	Arg	Val	Thr
	930					935					940				
Gly	Phe	Pro	Tyr	Pro	Ala	His	Phe	Ala	Lys	Leu	Glu	Val	Asp	Glu	Val
945					950					955					960
Glu	Thr	Arg	Arg	Thr	Thr	Leu	Arg	Cys	Pro	Ala	Asp	Val	Thr	His	Phe
				965					970						975

-continued

---

Leu	Asn	Gln	Arg	Tyr	Glu	Gly	His	Val	Met	Cys	Thr	Ser	Ser	Glu	Lys
			980					985						990	
Lys	Ser	Val	Ser	Gln	Glu	Met	Val	Ser	Gly	Ala	Ala	Ser	Ile	Asn	Pro
		995					1000						1005		
Val	Ser	Lys	Pro	Leu	Lys	Gly	Lys	Ile	Leu	Thr	Phe	Thr	Gln	Ser	
	1010					1015					1020				
Asp	Lys	Glu	Ala	Leu	Leu	Ser	Arg	Gly	Tyr	Ala	Asp	Val	His	Thr	
	1025					1030					1035				
Val	His	Glu	Val	Gln	Gly	Glu	Thr	Tyr	Ala	Asp	Val	Ser	Leu	Val	
	1040					1045					1050				
Arg	Leu	Thr	Pro	Thr	Pro	Val	Ser	Ile	Ile	Ala	Arg	Asp	Ser	Pro	
	1055					1060					1065				
His	Val	Leu	Val	Ser	Leu	Ser	Arg	His	Thr	Lys	Ser	Leu	Lys	Tyr	
	1070					1075					1080				
Tyr	Thr	Val	Val	Met	Asp	Pro	Leu	Val	Ser	Ile	Ile	Arg	Asp	Leu	
	1085					1090					1095				
Glu	Arg	Val	Ser	Ser	Tyr	Leu	Leu	Asp	Met	Tyr	Lys	Val	Asp	Ala	
	1100					1105					1110				
Gly	Thr	Gln	Tyr	Gln	Leu	Gln	Val	Asp	Ser	Val	Phe	Lys	Asn	Phe	
	1115					1120					1125				
Asn	Leu	Phe	Val	Ala	Ala	Pro	Lys	Thr	Gly	Asp	Ile	Ser	Asp	Met	
	1130					1135					1140				
Gln	Phe	Tyr	Tyr	Asp	Lys	Cys	Leu	Pro	Gly	Asn	Ser	Thr	Leu	Leu	
	1145					1150					1155				
Asn	Asn	Tyr	Asp	Ala	Val	Thr	Met	Lys	Leu	Thr	Asp	Ile	Ser	Leu	
	1160					1165					1170				
Asn	Val	Lys	Asp	Cys	Ile	Leu	Asp	Met	Ser	Lys	Ser	Val	Ala	Ala	
	1175					1180					1185				
Pro	Lys	Asp	Val	Lys	Pro	Thr	Leu	Ile	Pro	Met	Val	Arg	Thr	Ala	
	1190					1195					1200				
Ala	Glu	Met	Pro	Arg	Gln	Thr	Gly	Leu	Leu	Glu	Asn	Leu	Val	Ala	
	1205					1210					1215				
Met	Ile	Lys	Arg	Asn	Phe	Asn	Ser	Pro	Glu	Leu	Ser	Gly	Val	Val	
	1220					1225					1230				
Asp	Ile	Glu	Asn	Thr	Ala	Ser	Leu	Val	Val	Asp	Lys	Phe	Phe	Asp	
	1235					1240					1245				
Ser	Tyr	Leu	Leu	Lys	Glu	Lys	Arg	Lys	Pro	Asn	Lys	Asn	Phe	Ser	
	1250					1255					1260				
Leu	Phe	Ser	Arg	Glu	Ser	Leu	Asn	Arg	Trp	Ile	Ala	Lys	Gln	Glu	
	1265					1270					1275				
Gln	Val	Thr	Ile	Gly	Gln	Leu	Ala	Asp	Phe	Asp	Phe	Val	Asp	Leu	
	1280					1285					1290				
Pro	Ala	Val	Asp	Gln	Tyr	Arg	His	Met	Ile	Lys	Ala	Gln	Pro	Lys	
	1295					1300					1305				
Gln	Lys	Leu	Asp	Leu	Ser	Ile	Gln	Thr	Glu	Tyr	Pro	Ala	Leu	Gln	
	1310					1315					1320				
Thr	Ile	Val	Tyr	His	Ser	Lys	Lys	Ile	Asn	Ala	Ile	Phe	Gly	Pro	
	1325					1330					1335				
Leu	Phe	Ser	Glu	Leu	Thr	Arg	Gln	Leu	Leu	Asp	Ser	Ile	Asp	Ser	
	1340					1345					1350				
Ser	Arg	Phe	Leu	Phe	Phe	Thr	Arg	Lys	Thr	Pro	Ala	Gln	Ile	Glu	
	1355					1360					1365				
Asp	Phe	Phe	Gly	Asp	Leu	Asp	Ser	His	Val	Pro	Met	Asp	Val	Leu	
	1370					1375					1380				

-continued

Glu Leu Asp Val Ser Lys Tyr Asp Lys Ser Gln Asn Glu Phe His  
 1385 1390 1395  
 Cys Ala Val Glu Tyr Glu Ile Trp Arg Arg Leu Gly Leu Glu Asp  
 1400 1405 1410  
 Phe Leu Ala Glu Val Trp Lys Gln Gly His Arg Lys Thr Thr Leu  
 1415 1420 1425  
 Lys Asp Tyr Thr Ala Gly Ile Lys Thr Cys Leu Trp Tyr Gln Arg  
 1430 1435 1440  
 Lys Ser Gly Asp Val Thr Thr Phe Ile Gly Asn Thr Val Ile Ile  
 1445 1450 1455  
 Ala Ser Cys Leu Ala Ser Met Leu Pro Met Glu Lys Leu Ile Lys  
 1460 1465 1470  
 Gly Ala Phe Cys Gly Asp Asp Ser Leu Leu Tyr Phe Pro Lys Gly  
 1475 1480 1485  
 Cys Glu Tyr Pro Asp Ile Gln Gln Ala Ala Asn Leu Met Trp Asn  
 1490 1495 1500  
 Phe Glu Ala Lys Leu Phe Lys Lys Gln Tyr Gly Tyr Phe Cys Gly  
 1505 1510 1515  
 Arg Tyr Val Ile His His Asp Arg Gly Cys Ile Val Tyr Tyr Asp  
 1520 1525 1530  
 Pro Leu Lys Leu Ile Ser Lys Leu Gly Ala Lys His Ile Lys Asp  
 1535 1540 1545  
 Trp Asp His Leu Glu Glu Phe Arg Arg Ser Leu Cys Asp Val Ala  
 1550 1555 1560  
 Glu Ser Leu Asn Asn Cys Ala Tyr Tyr Thr Gln Leu Asp Asp Ala  
 1565 1570 1575  
 Val Gly Glu Val His Lys Thr Ala Pro Pro Gly Ser Phe Val Tyr  
 1580 1585 1590  
 Lys Ser Leu Val Lys Tyr Leu Ser Asp Lys Val Leu Phe Arg Ser  
 1595 1600 1605  
 Leu Phe Leu Asp Gly Ser Ser Cys  
 1610 1615

&lt;210&gt; SEQ ID NO 3

&lt;211&gt; LENGTH: 264

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Tomato mosaic virus

&lt;400&gt; SEQUENCE: 3

Met Ala Leu Val Val Lys Gly Lys Val Asn Ile Asn Glu Phe Ile Asp  
 1 5 10 15  
 Leu Ser Lys Ser Glu Lys Leu Leu Pro Ser Met Phe Thr Pro Val Lys  
 20 25 30  
 Ser Val Met Val Ser Lys Val Asp Lys Ile Met Val His Glu Asn Glu  
 35 40 45  
 Ser Leu Ser Glu Val Asn Leu Leu Lys Gly Val Lys Leu Ile Glu Gly  
 50 55 60  
 Gly Tyr Val Cys Leu Val Gly Leu Val Val Ser Gly Glu Trp Asn Leu  
 65 70 75 80  
 Pro Asp Asn Cys Arg Gly Gly Val Ser Val Cys Met Val Asp Lys Arg  
 85 90 95  
 Met Glu Arg Ala Asp Glu Ala Thr Leu Gly Ser Tyr Tyr Thr Ala Ala  
 100 105 110  
 Ala Lys Lys Arg Phe Gln Phe Lys Val Val Pro Asn Tyr Gly Ile Thr  
 115 120 125

-continued-

Thr Lys Asp Ala Glu Lys Asn Ile Trp Gln Val Leu Val Asn Ile Lys 130  
 135  
 140  
 145  
 Asn Val Lys Met Ser Ala Gly Tyr Cys Pro Leu Ser Leu Glu Phe Val 150  
 155  
 160  
 Ser Val Cys Ile Val Tyr Lys Asn Asn Ile Lys Leu Gly Leu Arg Glu 165  
 170  
 175  
 Lys Val Thr Ser Val Asn Asp Gly Gly Pro Met Glu Leu Ser Glu Glu 180  
 185  
 190  
 Val Val Asp Glu Phe Met Glu Asn Val Pro Met Ser Val Arg Leu Ala 195  
 200  
 205  
 Lys Phe Arg Thr Lys Ser Ser Lys Arg Gly Pro Lys Asn Asn Asn 210  
 215  
 220  
 Leu Gly Lys Gly Arg Ser Gly Gly Arg Pro Lys Pro Lys Ser Phe Asp 225  
 230  
 235  
 240  
 Glu Val Glu Lys Glu Phe Asp Asn Leu Ile Glu Asp Glu Ala Glu Thr 245  
 250  
 255  
 Ser Val Ala Asp Ser Asp Ser Tyr 260

<210> SEQ ID NO 4  
 <211> LENGTH: 2131  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 DNA

<400> SEQUENCE: 4

atggtcctag ctgttaagg taagtaaat atcaatgagt tcatcgatct gtcaaagctc 60  
 gagaacctc tcccgctgat gtcacaagct gtaaaagatg tcaagttgat 120  
 aagattatag tccatgaaaa tgaatcatg tctgaagttaa atcctctaagaa 180  
 ctatagaag gtgggtatgt tgcctatgct gctcctgctg tgcctggtgaa 240  
 ccagataat gccgtgtgtg tgtgagtggaa tctggccccta tctcctgtat 300  
 cactaatacc atcaaaagcc cgtgcccga gatattcccg tggcagagca 360  
 cactatgctg atgccatct ctataatgct cgaaccgctc tgcaaagttc 420  
 tcaaccctc tggtttggc atgatataga tgtagtcaag cttatcgtt 480  
 tctctctgat tccgttatacc tccgtaacct gcaaccgctgt tcaagttccc 540  
 acaaccgggg gaaacctggtgt gtttcaatag cccgggtagca tgaaccat 600  
 actcaccctg agaacagagaa agatctgtgc cgaagcggcg tcccaaccgt 660  
 cgtcctctga cgtttcaactg tgaagccgc cgtgtagtga cgcgccctca 720  
 gatcacacag gttccagcgc aaccgtgacac gacgtcctcc tccgttatac 780  
 cgaaccctct agaacaccat cgtctgatgc tttccttca cggccactctg 840  
 cccgtttcga tgcctgtatac gacagtttcc gctcaagcaa cagctcctcc 900  
 tccggttaaa acgtttggca tgatccag tgcataccatc aagttcctcc 960  
 gcagcgcacc taaccagaga taaccatag tctgcaagctc tccgatatac 1020  
 gtagaagca acatccgatatac aactctgcaag gacatctgagc gacatctc 1080  
 tctcctctag aatgacatca aactctgcaag gacatctgagc gacatctc 1140  
 tcaaccctcc tcccccggca ataaagggcc ataaagggcc gctgctatc 1200  
 tccaagcctc tctctgagaa gttcatttcc catcgtctg cccgtccatc 1260

atggtcctcag tctctaaagg taaggtcaat atcaatgagt tcatcagact gccaagctc  
60  
gagaaactc tcccgctgat gtcacaagct gtaaaagagtg tcatcgtctc aaggtctgac  
120  
aagattatctg tccatgaaaa tgatcactg cctgaaagtca atcctcctaaa aggtctgaaa  
180  
cctatgaaag gttgggtctatg tgcctcctagt ggtcctcctagt gctcctcctagt gctggaactta  
240  
ccagatcatc gccggtctggtg tgtcaggtctg tcttgccctca tctcctcctca cactcctgctat  
300  
cactcaaaccc atcaaaagccc cgtctgcccga gatctatcccga tggcctcctcga taaaccacagc  
360  
cactctgctcgt atctccactcgt tcaatctgctc gcaaaagctc tgcctcctcctc tctgcccctc  
420  
tcaacccccgt cgtgtctctcgt atgctatctgca tctctatcctc tccacagagc  
480  
tctctcctcgt tccgctcctcgt gcaaccgcccgt gttcctcctcgt tcaagctcccga  
540  
aacctccgctcgt gaaactcctcgt gttctcctcctcgt tccctcctcctcgt tccctcctcctcgt  
600  
actcctccctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
660  
cgtctcctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
720  
gactccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
780  
cggaaaccctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
840  
cccgctctctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
900  
cccgctctctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
960  
gcaagctccctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
1020  
ggagaaagcacc atctccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
1080  
tctcgtctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
1140

<400> SEQUENCE: 5

<210> SEQ ID NO 5  
<211> LENGTH: 2053  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

tctcttctac tgcctcctcgt cggcagcaga tccagagcca gaaacaacct gttcctcccgc  
1320  
ggcagcaga gtaagactc cctccctcga tgcctcctcgt cagaaataa actgctcctcgc  
1380  
tgcctcctcgt gctcctcctcgt caaccctcctcgt caaccctcctcgt caaccctcctcgt  
1440  
aaactcctcgt atctctcctcgt gctcctcctcgt cgtcctcctcgt tctcctcctcgt ctaagacactc  
1500  
aatcctcctcgt tctcccaatga ctagctcaaa aactagctatc agtaccctcctcgt cttatctcag  
1560  
tgatctcctcgt tgcctcctcgt agtaccctcctcgt agtaccctcctcgt agtaccctcctcgt  
1620  
gagatctcgt agagcagcagc aggcacacct ggggtctcctat taccactcctcgt cgtcctcaaaa  
1680  
gctcctcctcgt tccaactca cgtctatcga acaaaagctg cagaaagaaa  
1740  
cactcctcctcgt gctctcctcctcgt atctcaaaa atctcaaaaatg agtaccctcctcgt actgcccctc  
1800  
gctcctcctcgt tctcctcctcgt tctcctcctcgt tctcctcctcgt tctcctcctcgt  
1860  
ggagaaagta acgagctcctcgt acccactcctcgt cctcctcctcgt aagtctcctcgt  
1920  
tggctcctcgt gaggactcctcgt caatgctcctcgt taccctcctcgt agtaccctcctcgt  
1980  
aaaagagctc cggaaataatc aggtcagggg cgtcctcctcgt gaaagcctca  
2040  
aacaaaaagta tctcctcctcgt tctcctcctcgt gttcctcctcgt tctcctcctcgt  
2100  
gagctcctcgt agtaccctcctcgt atctcctcctcgt  
2131

-continued-

1200 ttaactccca tccccgggca ataaaggcgc gtcgcctacc cactttcttg cccgctccata  
1260 tcaacggct cctttgagga gttcatttcc catcgtttcc tgcctggagca ggcgctggag  
1320 tcttataac tgcctcattg cgcctcattg tccagagca ggaacaacct gttcctccggc  
1380 ggcgacagca gtaagacttc cctccctgta tgccttacc cagaaataa actgctggc  
1440 tgcctaccca gttcgcggg caacgaggg gaccgctcacc ccgggtccaa agctcctgctg  
1500 aacaattcgc atcttttctc gttggtgta cgcctctgct tctcctggc ctaagacatc  
1560 aatcactcct tctccaatg tgaagtctc catggtctg cacgaaatg aagaaagcggaa  
1620 cgaagccaca ctgggtccat atcacctgc tgcctgctaaa agcggttcc agtttaagt  
1680 ggtccccaat tacgttatca caacaaggaa tgcgaaagaa acatatactg aggtccttagt  
1740 aatatataaa aatgtaaaaa tgaagtgcggg ctacctgccc tgcctcattag aatttctgctc  
1800 tgtgtgtaat gttatataaa attatataaa atttgggtctg agggagaaag taacgaggt  
1860 gaaagatgaa ggaacctcggaa agaaagtgtt gatgagttca tggagaaagt  
1920 tccaatgctc gttagactcg caaagtctc aaccataacc tcaaaaagag gtcctgaaaaa  
1980 taataataat taggttaagg ggcgtccagg cggaaagcct aacccaataa gttttgtaga  
2040 agtgaaaaaa gagttagata attttagata agatgaagc gagagcgtccgg tccggtgattc  
2053 tgaatcgtat taa

<210> SEQ ID NO 6  
<211> LENGTH: 2128  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

60 atgctctcag ttgtatataa cgttccagc acagcaatc aagagatcgc ctgatggtat  
120 cgggtgtagag gtcgcaagaa atcacatgaa cgcaggtgaa cgcaggtgaa ggttccgagt  
180 taacgctcgc tccgcccagt ggcgagaaat atccctgctg accctgctgaa cgggtatccg  
240 gttcgtctgg caatacctcaac atcaaccacgc ttgggtggtt ttgtcagcgcg gctatcagct  
300 cttcaatcgc cgttaagctgc gcttgcctgag ttctcccccgtc gaccctgccc tccatcagca  
360 gttctctcctg cttgttctgc gcttccgaaac caatgcccata agagaggtta aagccggacag  
420 cagcagttcc atcacaatcacc acgattcccat gttcattctgc ccagttccagc atctctccag  
480 cgtkaaggtta atctcagaggt cggtaaggtat cgggtcccccatt caggtcccatt aatgctggt  
540 cgtgctcccaat cagcacaagctt taatcagaa cgtttctggtt taatcagaa cgtttctgccc  
600 caaagccagct aagatgagaa agtttctggtt taatcagaa cgtttctgccc tccatcagccc  
660 ctgacccggat gcccagagccag agctcaggtatgaa taccacaacctc tgcctcggctt ttggtcgtgaa  
720 ccgcaacgttcc atagagataa ccttccaccgc gttctgcccagg gttctgctggatc accacttctgca  
780 aagttcccctc agttgcccctt gttccagttccg caaacctggttg atccctgctg atccagttccaaa  
840 cgcctcagccat caccatctgg ccccccagc agttccaaagaa cgcctcagc agttcctcctgctg  
900 ccgcaatgctc caaccacgggt ataatcgttccca ccaaggtggtt cggctcgttgggtg tagagcattca  
960 cgcctcctctc gttatctcctg atctcctccg caatctcaaaa gattcagctgca ttttctcctcctg  
1020 cgttctctctc gttatctcctg atctcctccg caatctcaaaa gattcagctgca ttttctcctcctg  
1080 caacaacgggt gatcaatcagca ctttcccccgg caatcaaacata cggctcgttggaca tccgcttccaaa  
1140 atcgtctatata gcccctcctca tgcctcctca cttcctcctca atctcctgata cttcctcctcgt

-continued-

aatgagtgaac cgcatacgaac cgcagcagca taagcttgcc tcgccaacct tcctgkatabaa  
1200 agacttcctcg ctagatataccag acgtctgcccgc catabatccgc aatcctccgaact  
1260 agatcgttaaa actgcccctggc acagcgaatg cccggccttc tcgtaacccgc ccttcccaac  
1320 aacgctgctac aatccccaag tttctcgcgat aggttaaggtaa aatcatabatg agttatccga  
1380 aacgctcaagc tctgcaagaaac tctcccgcgc gatggttccacg cctgkatabatggt  
1440 tctgcaaggtc gatbaagatcaa tggkccatga aatgacatcaa tgcctccgaaq taakatcctc  
1500 tccaaggtc gatbaagatcaa tggkccatga aatgacatcaa tgcctccgaaq taakatcctc  
1560 aaaaaggttca aaacctatag aaggtgggtca tgttctgctta gkctgctctg tcgtgkccgg  
1620 tggatgggaat ttaaccagatca actgcccggg tgggtgktagat gkctccatgg tcgacaagag  
1680 aatggaaagaa ggcggacggaaq ccaaccctggg gkccatcatca actgctccgctg ctaaaaagc  
1740 gkctcagttc aaggtggkcc caaatccaggt tatabccaaca aaggtatccag aaaaagaaact  
1800 atgacaggtc ttagtaataca taaaatgt aaaaatgagt gccccttgc actgctccatg  
1860 attagaatc gkctccctgct gkctatgkttca taaaatcaat aataatctgg gkcttgaagga  
1920 gaaagtaaccg agtggkgaacg atggaggaacc catggaaact tcggaagagag tggktagtaga  
1980 gkctcagtag aatgkctcaaa tgcctggktag actcgcgcaag tttcccaaaa aatccctcaaa  
2040 aagaggtccg aaaaatataca ataattagg taaggggcgt tcagggcggaa ggcctcaaac  
2100 aaaaagttc gatgaagctg aaaaaggtt tgaatattg attgaagatg aagcggagac  
2128 gkccggkccg gatcctgact cgtatcaa

<210> SEQ ID NO 7  
<211> LENGTH: 2128  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA  
<400> SEQUENCE: 7

atgctctag tggtaaggt taagtaaat attaatgagt ttaactgatct gkcaagctc  
60 gaaacttc tcccgtcgat gkccacgcct gktaagaggtg tcaatgkcttc aaggttagat  
120 aagatattag tccatgaaaa tcaaggttc aggcacaagca catcaaaagag atccgctgagt  
180 gkctcgggtc gacgctccgca gaacattaca tgaaccgaggt gkctcgggtccg  
240 aattatccag tgcctctctaa tccctgktaa gkctcctctc tccgctctccg cctcctccgctg  
420 taccgtctctc tccgctctctc gcccctctc aaacccaatg ctaaaagagag gktaaaagccg  
480 taaagcagcag ttcctatcaat cacaaccagatg ccatgkctcat ctgcccacagtc gagcctatcctc  
540 tcaagcttaag gktaatggcga gktaacggtag gkctcggccc caatccacagtc cacttaatgctg  
600 tggkctccgtag caacatcagcaca gktaacgtag cctctctgacat gkcaagctccgc atcctccatga  
660 cgaaccaaaag cagtkaaaatga gaacggkttg tggktaatca ggaactgkctc gcccctcaact  
720 gccaactgccc gkctatccgcaac ggcgaagcggg tagatatacaac actcctgctg gctctctgct  
780 gktaacggca gkctcattagat ataaactca cccggkttgccc aggaagttgctgg atccacaact  
840 tgaaaagctc cgcctcagttcc tggkccagtt gcaaacccaact gkctgkattccggc atccacggcagtt  
900 tcaaacgctga catcaaccatc ggcccaaccatc ggcacagctcaaa cagaaagcggctg gktaacagctc  
960 tgcgcaacat gkctcacaact gkctcacaact gkctgkattccg gkctcacaactg  
1020

-continued-

1080 attaacgctgc gattggattcc ggcataagta aagaaatcat ggaagttaaga ctgcctttcc  
1140 ttcgccgtttc gtcctggtatt ctcctggtatg tccgccattcc caccattccc ggcgggatatg tccgcccagt cagtccgttg  
1200 tccaccacaaga cggctgatatag taccatttcc ccggcacaataa catacggcgtt gacatccgctc  
1260 tcaatggcgg tatabgcggcc ctgattgcacc atccacttccc gattattatga cccaacttgg  
1320 ccgtaatgag tgaaccgcatc gaaagcagc aggatagctt ggccttgcaca acccttccgtt  
1380 atcaaatgactt cgcctgctgatat ccaagacttgg cccgcataatc taccgaaatcc tgcactccgctg  
1440 aactgatctgt taaaacctgccc tggcaccagca attgcaccggc tttcttctgtaa cgcgccttccc  
1500 caccacaagc gatcaattcc acagtttccg cgtatgataca ttgtcctgagc taatcctctc  
1560 aaaaagttgta aaacctatag aagttgggtat tgttctgcta gttcgtcttg ttgtgtccgg  
1620 tggattggaaat ttaaccagata atcgcctggg ttggttggatg ttccctcattgg ttgacacaagag  
1680 atggaagaag gcggaacgagc ccaaccctggg gtcataatca accctgcttg ctaaaaagcg  
1740 gttcctcagtt aagatgtgctcc caatatcaggg tatatacaca aaggtatggcag aagaagaaatc  
1800 atcggcagctc ttagtataatc taaataatgt aaaaatggat gcgggtctcat gcccttctgc  
1860 attagaatctt gttgctcctgtgt gtaattgttca taaataatca ataaatctgg gttttagaggg  
1920 gaaagtaaacg agtgtgaaacg atggaagacc catgaaatc tccggaagag ttgtttagatga  
1980 gttcattggag aatgttccaa ttctcgttag accctgcacaag ttcccaaca aatccctcaaa  
2040 aagaggtccc aaaaatataa ataattagg taaggggcgt tcaaggcggaa ggcctcaaac  
2100 aaaaatctt gatgaagttg aaaaagattt tgaataatgg atggaagatg aagcggagac  
2128 gtcgggtccg gattcctgatt cgtatca

<210> SEQ ID NO 8

<211> LENGTH: 2138

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

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

<400> SEQUENCE: 8

60 atgctctcag ttgtaagaag taagttaaat attaatgagt taaatcgatct gccaagctc  
120 gaaaatctc tcccgctgat gttcaacgct gtaaaagatgg tcaatggttcc aaaaggttgat  
180 aagatattag tccatgaaaa tgaatcattg cttgaaatga atctcctcaaa aggttctcaaa  
240 cctatatgaag gttgggtatgt ttgcctttagt gttcctttagt ttgctcggttga gttggaatca  
300 ccagatcaat tcccggttgg ttgtcagttaac gttcctcagcca agtgcacatca aagatgattcc  
360 ttgatgttacc gttgttggacg tccagaaaca tcaatctgac gcaagttgatc ggaacgctcgg  
420 ggttcggattt accgcttctgt ctcgcagttg gctgcgaataa tccctgctgaca ccttgcggac  
480 ggttatcattc cctcgtttgg atatacaca tcaaccacgc ttgggttggttt ttgttccacgctg  
540 ctatcagctc tttaatccgc ttgtaagttcgt cttgctttagt ttccctcctg accctcctct  
600 cgtctgtaacg tctcttctcggc ttgttctcggc cttcgaaaaacc aatgctcaaa gagtagttcaaa  
660 aagcccgacag agcaattcca tcaatcaca tcaatcctcctc cagtctcagagc  
720 tctctcagc gttaaagttgaa ttgcagttgac ggttaagttgac ttgcctcccaatc cagtccatca  
780 atcctgctgctc gttcctcaccac atccagtttat cgaatccctt ctccacagc aagctcctct  
840 caatcagcaac aagagttca aagtatgaaacg gtttcttggttt aatcagttgaa cctcctccct  
900 tccactctgccc atgacccgggatg cggacccgcaag gctcagctc atcaaccactc gttcctgctctt  
960 ttgctctgctg acagcttcca atagatatac cctcaccacc cttcagctatca ttgctgattcca

60 atgctctcag tctgttaagag taagtaaat atlaatagat tlatcgtactt gtcacaagct  
 120 gaaacactc tcccgctgat gtcacacgct gtaaaagagtg tcaagtgcttc aaaagtgat  
 180 aagatctatgg tccatgaaaa tgaatcattg tctgaagatba atctcctcaaa aggtgtcaaaa  
 240 cctatagaag gctgggtatgt tgcctcctgctg gctcctgctg tgcctcctgctg gctggaatct  
 300 ccagatctatc gccgctgctgctg tctgagtgctg tgcctcctgctg tgcctcctgctg gcaagagctg  
 360 gacgaagcca cactgctgctc atactacact gctgctgctc aaaaagctgt tcaagtctaaa  
 420 gctgtcccca atctacagctat tacaacaag gattgcagaaa agaacatctag gcaagctctca  
 480 gtaaatctca aaaaatctaaa aatgagtgctg gctcctcctc cttctgctct agaatctgctg  
 540 tctgtctgta tctgttctaaa aatatataata aatctgctg tgaagtgagaa agtcaacgagt  
 600 gtaaacagatg gaggacccat ggaactctg gaagagctg gaaagagctg tctgatactg catgagaaat  
 660 gttcccaatg ccgctctgaact taaagtgctca ggcacagctc atcaaaagagaa tccgctcgtatg  
 720 tctcctgctg agctcctcag aacatctaat tgaacagagtg gatcctgagctg tccgctcgtatg  
 780 gttctacagct tgcctcctcag aatctcctcga aatctcctcga aatctcctcga ggaacagctat  
 840 ccgctctcct ggcacatctc ggcacatctc ggcacatctc ggcacatctc ggcacatctc ggcacatctc

<400 > SEQUENCE : 9  
 <210 > SEQ ID NO 9  
 <211 > LENGTH: 2128  
 <212 > TYPE: DNA  
 <213 > ORGANISM: Artificial Sequence  
 <220 > FEATURE:  
 <223 > OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA

1020 ccaattgcaaa agtccctgctc cagctgcaac cactctgta tccgcatca  
 1080 gcaagtccaac gctgacatca ccaatgcca ccaatgcca gctggtatc  
 1140 agtctctgctg gacatgctc accaagtgaa tctcctcga ccaggtgtctc ggcgtggtg  
 1200 agagcctatc gctcctgctg atctcctgctg agtcaagaa atctatgagaa taagactgct  
 1260 tctctctctc gctctctcctg gtaatcaca tctcctcctc gattcctcctc cagtctcagct  
 1320 cgtctgctcga caaaagtg atactcaca tctctcctcctc ataacatctc ggcgtgcat  
 1380 cgtctcctaaa tgcctctatg cctcctcctc gctcctcctc tctcctctat tgaaccca  
 1440 cctctgctca atgagtgacc gcatcgaaac gcaagcagat agctctgctc gcccacact  
 1500 tctgtataaa gactctcctc tgaatcagaa cgtctcctcctc ataatctcga atactctgcat  
 1560 cgtcctcctc atcgttaaaa cctcctcctc cagcaatctc cctcctcctc tgaacgctc  
 1620 tctcccaaca agctctgctc atctcctcga atctcctcctc tgaagtgctc catgctgctg  
 1680 aagagatctg aagagctcga cgaagccaac cctggtctc atcaacctc tgcctcctcaaa  
 1740 aagcgtcttc agttaaagt gctcctcaat ttaggtatca caacaagaa tgcagaaag  
 1800 aacatctatc agtctctatg aatatataaa atgtcaaaa tgaagtgctc ctaactcctc  
 1860 tctcctcctc aatctctcctc tctctctcctc gctctcaaaa ataatcaaaa atctggtctg  
 1920 aggtagaagaa taacagatgt gaacagatgaa ggaacccatg gaaactctcga agaaagtgt  
 1980 gatgagctca tgaagatctg tccaatctc gctgaactc gaaagtctc aaccaatcc  
 2040 tcaaaagag gtcgaaaaa taatatatct taggtaaag gctcctcag cgaagcct  
 2100 aaaaacaaaa gctctgctga agtctgataa agtctgataa atctgctga agatgaagcc  
 2133 ggaagctcgg tccggtctc tgaatctat taa

60 atgctctcag tctgttaagag taagttaagt ttaactgactc gtcacaagctc  
120 gaaacatct ccccgctcagat gttcaagagctg tcaatgcttcc aaaagctgcat  
180 aagatctatgg tccatgaaaa tgcatactatg tctgaaagctaa atcctcctaa  
240 ccttatgaaag gctgggtatctg tctgctctatgct gctcctctgctg atgctgaa  
300 ccagatcaatc gccggtgctgctg tctgctgctgctg accaagagcaat ggaagagc  
360 gacgaagcca cactgctggctc atactcaactc gctgctgctgctc aaaaagcgtc  
420 gctgctcccaa atctacagctat tacaacaaga gctgctcagaaa agaaactatg  
480 gtaaatctatc aaaaatctcaaa aatgagctgctg gctcctcactcctc agaa  
540 tctgctgctgca tctgtttctatc aaatctatctc aaatctatctc tgaagctg  
600 gctgaaagcagctg gaggccatcctg gaaactctcctg gaaagagctg cactg  
660 gctcctcactc cgtctctgagctc cgtcaaaagctc cgtcaaaagctc agctc  
720 aatctaatctatc atctcagctcag gggctgctcag ggcctggaagcc ctaaa  
780 aaggtctgaaa aagagctctgaa taactctgact gaaagctgaaag ccgagctc

<400> SEQUENCE: 10

<210> SEQ ID NO 10  
<211> LENGTH: 2128  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

900 gctcctctaat cgcctctgag tgcctctgct gaaactctccccc gctgactcctc  
960 aagctctctc cgcctctgctg cccgctctcga aaaccaatgccc taaagagagc  
1020 cagcagcagct tctatcaatcc acccaagatgc catgttctatc tgcctcagctc  
1080 cagctaaagag gtaactgctagag aatctgctccccc aatccagctcc attaatc  
1140 gctcctctgctc catcagcagctc tctatcagatc cttctgcccagc caagtcc  
1200 gaccaaaagcc agttaaagctag aaccgtttctg ggttaaatcag gaaactgctc  
1260 cccatctcagc gctgcccagcag cgaagcctgctg agatctatcaca cctcctg  
1320 tgaagcaccag tctatgagaa taaacctcacc cctgctctgccc aggtctgccc  
1380 gcaaaagctccc gctagctgctcct tctcagctgctg caaacaccctg tga  
1440 caacgctcagc atcaccaccatc gccaccaccatc gccagctcaaac agacgctg  
1500 gcgctgacatg cgtcacccagc gctgatatctc ccaaccaggt gctcctgctg  
1560 tcaagcctcag atctgctcctcag gctatgctcag agaaatctatg tgcctctc  
1620 tgcctctctc gctcctgctcagc accatctcccg gctggtgctatgct cctg  
1680 tcaaacaccag gctgctctgctc cagacgctgctc cgcatactatc agaa  
1740 ccaatctgctc atagcctcctc tcaatctcctc atctatctcctc ccaactg  
1800 cgtatctgagct gaccctcagctc aaacgctcagca cgtatcagctc cctc  
1860 taagacatctc gctgctctgctc cagacgctgctc cgcatactatc agaa  
1920 actgctcagctc aaaaactcctc ggcaccagcaaa tctgctcctcctc gctc  
1980 accaagcctc atcaatctccc cagctctcctc gctcctcagcaag tctcga  
2040 aagagctcagc aaaaatctatc atcaatctagc taagggctcctc cagctcga  
2100 aaaaagctctc gctgagctcag aaaaagagctc tgcatactctg atctgaa  
2128 gctcctctgctg gctcctctgctc gctatctgctc cgtatctaa

-continued-

60 atgctctcag tctgttaagag taagttaaat attaatgagt ttaactgatct gccaagctc  
 120 gaaacactc tcccgctgat gctcaacgct gtaaaagagtg taaagtgcttc aagtgctgat  
 180 aagatctatgg tccatgaaaa tgcatactgg tctgaaagta atctctcctaaa aggtgtcaaaa  
 240 cctatagaag gctgggtatctg tgcctctgctg ggtcctctgctg tgcctctgctg gctggaatctta  
 300 ccagatcaatc gccgctggctgg tctgagtgctg tgcatactgg tgcatactgg tgcatactgg  
 360 gacgaagcca cactgggctc atactcaact gctgctgctgctg aaaaagcgtt tcaagttaaaa  
 420 gctgtcccaaa atcaactgat tacaacaaga gatgcagaaa agaaacatag gcaagtctta  
 480 gtaaatatct aaaaatctgaaa aatgagtgctg ggcctacatgccc cttctgctcct agaatctgctg  
 540 tctgtctgtca tctgtttatca aatatatata aatatgctt tgaaggagaa agtcaacgagt  
 600 tctgaaacgagtg agggaccat ggaactctcg gaagagtg tctgatatgctt catgaggaat  
 660 gttcccaatg tggctagact cgtcaaatct cgtcaaatct cgtcaaatct cgtcaaatct

<400> SEQUENCE: 11

<210> SEQ ID NO 11  
 <211> LENGTH: 2134  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 DNA

840 tctgatctca aggtctcag cacaagcaat caaagagatc gctgatgta tctggtgtag  
 900 cgtccagcaaa catcaatctg accgaagtgat tccgaacgctc agggatcagat tcaagcgtg  
 960 cttcccgcaag tggcggcaaa tctcccgctg cacccttgctg acccttgctg aggtgtatcc ggtctgctg  
 1020 caatctcaca cactcaacag cctgggtggt tttgtcagc ggtctcagc tcttcaatcag  
 1080 cctgttaagtg cgtctctctga gttctcccgct tgaactcctc tctcctctctc agttctctctc  
 1140 gctctgtctg ccctctcgaaa ccaatgctca aagagaggtc aagaccgaca gcaagcagtt  
 1200 catcaatctc cactgatgcca tgtctatctg ccaagtctg aag catctctca gcttaaggt  
 1260 aatgctaggt accggtaggat tggccccaa tccagttccat taatgctgctg tctgctgcaaca  
 1320 tcaagcagtt atcgaatcct tgcaccagca agtccgcatc tctcatgacga ccaagccag  
 1380 taaagttagaa cgttttgtggt tcaatcagga actgtctgccc cttcaactgc actgtaacgga  
 1440 tgcagcaag aagcgggtag atactcaact ctgtctgctc tctggctctg acccaagct  
 1500 catagagata accctcaacc ggttgccaga ggttgccgact ccaactctg caagctccgc  
 1560 tagtctctctg tccagtctga accactgtt gatccgcatc acgagttca acgctgacat  
 1620 caccatctg caccaccctg cagtcaacaag accgctgctg acagttctg gcaacatgctg  
 1680 tcaaccaggt gatctctcc acccaaggtc tccgctgctg gtagagcatt acgctgctg  
 1740 ggtctccggc atagttaaag aatcaatgga agtlaagactg cttctctctg cgttctctg  
 1800 cgttatctca cactcccgct ggtatagctt gccagttcag tctggtctc accaacaag  
 1860 tgatagctca acttttcccg gcaatcaaat accgctgctg acgctctca aatgctgctat  
 1920 agcccgctctg atgtctccatc actctctgat tatgaccca cacttctgctg taatgagtg  
 1980 ccgatctgaa accgacagct atactgctg cttgccaaac ttcgctata aagactctg  
 2040 gctgatcaaca gactctgccc gatcaatcac gaatctctg atcggcgaac tgatctgtaa  
 2100 aactgctcag cacagcaat gcccgctt cttgtaacgc gctctccca ccaagctgat  
 2128 caatccaca gttctgctga tgtatca

600 tgcatacgtcac accctttccccc gcaataaacat acggtcgtgac acatgcttca atcgtcgtat  
540 cgttatatcac cacttcccggc gggatagtc gccagttcag tctgcttcc accaacaacgg  
480 ggtatccggc atagtttaag aatccatgga agtcaagactg cttttcttgg ccgtcttccgt  
420 tcaaccaggt gatcatctcc acccaggtgt tgcgctggtg ctgagagcat acgctcgtgat  
360 caccatggc caccaccctgc cagtcaaacg acgctggtgt acagttcttg gcgacatggc  
300 ccagtatatc gccgtggtggt tccaccctgt gatcccgcatc acgctcagcca acgctcagcat  
240 cttatagaaq gttgggtatgt tctcctgctg gttcctgctg tgcctcgtgaa gttggaattca  
180 aagattatgg tccatgaaat tgatcatg tctgaaatca atctcctcaaa aggtgtcaaaa  
120 gaaacatct tcccgctgat gttcacagct gtaaaagagtg tcatggttcc aaaagttgat  
60 atgctctcag tctgttaagq taagtaaat attaatgagt tcatcgtact gtcaaagct

<400> SEQUENCE: 12

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA  
<220> FEATURE:  
<213> ORGANISM: Artificial Sequence  
<212> TYPE: DNA  
<211> LENGTH: 1395  
<210> SEQ ID NO 12

2134 caccacagct gatcaatcc accagttcgg cgat  
2100 aactgtagtc taaaactgcc tggcaccgga atctgcccggc tttctgttaa cgcgcttcc  
2040 atcaagactc cgcctcgtaca ccagacgttg ccgcataatc ttagaatc tgcactcggc  
1980 ccgtatagag tgaaccgcagc gaacgcagc agtatagct ggcctgccc aaccttcggt  
1920 tcaatggcgg tatagcgcgc ctgtagctcc atcaactcct gatctatgac ccaacttg  
1860 tccaacaaca cgtgtagtag taaccttcc cggcaatca catcagcgt gatcactcggc  
1800 tctcgttctc cgtcgttatc caccatccc gtcgggtatg tctcgcagtt cagttcgttg  
1740 attacgtcgc gatgtagtcc ggcataatcc aagaaatcat ggaataagaa ctgcttcttc  
1680 tgcgcagcat gctcaccacc ggtgtagtag tccaccaccagg tctcgcgctg gtttagtagc  
1620 tcaacgtcga catcacacc tgcaccacc cagacgtcca cagacgcgtg gttacagctc  
1560 tgaagctcc cgtcagttgc tctcagtt gcaaccaccct gttgactcgc atcacgcagt  
1500 gtcgaccaca gttcatagag ataacctca ccggttgc accaggtcgc atccccact  
1440 gccactgacc gtagtgcgac gcgaagcggg tagtatcac acctctgctg gcttcttggc  
1380 ccgacaagac cagttaagta gaacgttgg tggtaatca ggaactgtgc gcccttccact  
1320 tggctcgtga ccatcagcac gttatcagat ccttctggcc gcaagtccgc atctcctaga  
1260 tcaagctcaag gttatgtaga gttacggtag gtagtctggccc caatccagtc cattaatggc  
1200 acagcagcag ttcactcatc caccacagtag ccatgctcat ctgcccaccatc gagctactct  
1140 taagttctc tccgctctgt gcccgctc gaaaccactg ctaagtagag gttcaaacgg  
1080 agctcttcaa tccctcgttaa gtcgcttgc tgaattctcc cgtctgacct gcctcctcgt  
1020 tccggtctgt tgcataatca ccaatccacc acgcttgggt gtttcttctc accgctctac  
960 agttacagc tctcctccgc cagtggcggc aatatctccc gttgcaacctg cggacgggta  
900 gttacgtggt gagcgtcgcg gaacattaca ttgaccagag tgatccggtag cgtcgggtcgt  
840 tctgattcgt attaactcgg ataacggtc aggcaccagca catcaaaagag atcgtctgatg  
780 gaagttagaa aagatttga taattgat gaaatgaaq ccgagactgc gttcgcggat  
720 aatataatca attatagtaa gggcgttca ggcggaaagc ctaaaccaaa aagttctgat

- cont Inued

atgctctcag tctgtaagag taagttaaat attaatgagt ttaactgactt gtcacaagctc  
 60 gtaaatctc tcccgctgat gttcaagagtg gttcaagagtg tcaatgagtg gttcaagagtg  
 120 gaaatctc tcccgctgat gttcaagagtg gttcaagagtg tcaatgagtg gttcaagagtg  
 180 aagatctcag tccatgaaat tgaatcattg tctgaaatgaa atctctctaa aggtctgtaaaa  
 240 cctatgaaag gctgggtatgt tctgcttctg tctgcttctg tctgcttctg tctgcttctg  
 300 ccaatctcag gctgggtatgt tcaatctcag atctgctcag atctgctcag atctgctcag  
 360 agcctcag atctcctcag atctcctcag atctcctcag atctcctcag atctcctcag  
 420 cctcctcag atctcctcag atctcctcag atctcctcag atctcctcag atctcctcag  
 480 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 540 aactgctcag cactgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 600 caatctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 660 gaaatctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 720 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 780 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 840 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 900 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 960 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1020 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1080 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1140 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1200 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1260 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1320 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1380 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1395 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag

<400> SEQUENCE: 13

<210> SEQ ID NO 13  
 <211> LENGTH: 1095  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 DNA

atgctctcag atctcctcag tctgtaagag taagttaaat attaatgagt ttaactgactt gtcacaagctc  
 60 gtaaatctc tcccgctgat gttcaagagtg gttcaagagtg tcaatgagtg gttcaagagtg  
 120 gaaatctc tcccgctgat gttcaagagtg gttcaagagtg tcaatgagtg gttcaagagtg  
 180 aagatctcag tccatgaaat tgaatcattg tctgaaatgaa atctctctaa aggtctgtaaaa  
 240 cctatgaaag gctgggtatgt tctgcttctg tctgcttctg tctgcttctg tctgcttctg  
 300 ccaatctcag gctgggtatgt tcaatctcag atctgctcag atctgctcag atctgctcag  
 360 agcctcag atctcctcag atctcctcag atctcctcag atctcctcag atctcctcag  
 420 cctcctcag atctcctcag atctcctcag atctcctcag atctcctcag atctcctcag  
 480 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 540 aactgctcag cactgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 600 caatctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 660 gaaatctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 720 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 780 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 840 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 900 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 960 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1020 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1080 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1140 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1200 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1260 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1320 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1380 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag  
 1395 gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag gctgctcag

atggtctctag tctgttaagg taagttaat attaatgagt tcatcgatct gtcaagctc  
60 gagaacttc tcccgctgat gttcacgct gtaagagtg tcatggttc aaagtttgat  
120 aagattatg tccatgaaaa tgaatcatg tctgaaagtaa atctcttaaa aggtgttaaaa  
180 cttatgaaag gtgggtatgt tgcctctgtg tgcctcgtgaa gtgaaattca  
240 ccagtatatt gccgtgtgtg tcatgttt gccctccctg tgcgttttt cacogaagt  
300 catgcccagtc cagcttttt gacgcaaaaa agcccgccga ctcggttttg ggtcgcgagt  
360 gaagatctc tctctgttac ccgccaagcg caatatgctc tgcgaggttc caaaatccg  
420 gaattctcat accctgttca ccgacgaagc gctggaagcga tcaaaagagc ggtgtatcaat  
480 atccagccat gccaatgat actctcat ccaatgtctg gtgtatcatg agtgcagccc  
540 ggtcaagta tccaagcctg atcgcgtgat gatatacggc tgcatacagtt tctccctcca  
600 ggcocgaagta tctttttcca gtaacttctc gtaactgtg tgcgtgtca tttgaaatca  
660 aatatatca aatttgggt tgaagagaa agtacaagat gtaacagat gtagaccat  
720 gaaactctc gaagaagtt tgaatgagt catgagaaat gttccaatgt cgttatgact  
780 cgaagttc cgaaccaat cccaagaag agttccgaaa aatataatca attatgtaaa  
840 ggggcgtca ggcgaagc ctaaccaaa agttttgat gaagtgtgaa aagagtttga  
895 taatttgat gaagatgaa ggcagagctc ggtcgcgagt tctgattcgt attaa

<210> SEQ ID NO 15  
<211> LENGTH: 2399  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

atggtctctag tctgttaagg taagttaat attaatgagt tcatcgatct gtcaagctc  
60 gagaacttc tcccgctgat gttcacgct gtaagagtg tcatggttc aaagtttgat  
120 aagattatg tccatgaaaa tgaatcatg tctgaaagtaa atctcttaaa aggtgttaaaa  
180 cttatgaaag gtgggtatgt tgcctctgtg tgcctcgtgaa gtgaaattca  
240 ccagtatatt gccgtgtgtg tcatgttt gccctccctg tgcgttttt cacogaagt  
300 catgcccagtc cagcttttt gacgcaaaaa agcccgccga ctcggttttg ggtcgcgagt  
360 gaagatctc tctctgttac ccgccaagcg caatatgctc tgcgaggttc caaaatccg  
420 gaattctcat accctgttca ccgacgaagc gctggaagcga tcaaaagagc ggtgtatcaat  
480 atccagccat gccaatgat actctcat ccaatgtctg gtgtatcatg agtgcagccc  
540 ggtcaagta tccaagcctg atcgcgtgat gatatacggc tgcatacagtt tctccctcca  
600 ggcocgaagta tctttttcca gtaacttctc gtaactgtg tgcgtgtca tttgaaatca  
660 aatatatca aatttgggt tgaagagaa agtacaagat gtaacagat gtagaccat  
720 gaaactctc gaagaagtt tgaatgagt catgagaaat gttccaatgt cgttatgact  
780 cgaagttc cgaaccaat cccaagaag agttccgaaa aatataatca attatgtaaa  
840 ggggcgtca ggcgaagc ctaaccaaa agttttgat gaagtgtgaa aagagtttga  
895 taatttgat gaagatgaa ggcagagctc ggtcgcgagt tctgattcgt attaa

<210> SEQ ID NO 14  
<211> LENGTH: 895  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

-continued-

600 aactgcccctgg cacagcgaact gcccggctctt cctgtcaacgc cctctccca ccaacgctgact  
540 gctgtatatacca gacgtctgcc cgcataatcac gatatatctgc atcggcggaac tgbatcgttaa  
480 cccgcatcgaag acgacgacag ataacgctgg cctgcccaca cctcctgtata aagacctcctgc  
420 aagcccctcctg atctcctccatc actcctcctgat tctctgaccca cactctcctg taaatgagtgga  
360 tgbatcgtcac actcttcccg gcaataacat acgycctgtgac atcggctctca aatcggcctgat  
300 cggatcaatcac catctcccggc gggatcagct gccagctcacg tctcgtcttcc acacaacagg  
240 ggtatcccggc atatgctaaag aaatcctatgga agtcaagacctg cctctctcctg ccgtctctctg  
180 aagatctatgg tccatgaaaa tcccaggctgt cctcaggctgt gtagagcact acgctcctgat  
120 gaaacactc tcccgtcctg at gctcaacgct gtaaaagctg tcaatgcttcc aaggtctgat  
60 atgctctcctag tctgttaaagg taaagtcaat atcaatgact tcatcctgat gtcacagctc

<400 > SEQUENCE: 16  
<210 > SEQ ID NO 16  
<211 > LENGTH: 796  
<212 > TYPE: DNA  
<213 > ORGANISM: Artificial Sequence  
<220 > FEATURE:  
<223 > OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
DNA

960 cgcctcctcctg tgcctcctcctg gactcttccc gttctgacctgc tctcctcctgc acagctctctt  
1020 cggcctctcctg cccgctctcga aacccaatgcc taaagagaggg taaagccgga cagcagcagt  
1080 tctcatccatcc accaagctatgc catgctctcatc tgcaccagctc agcatctctct cagcgttaagg  
1140 gtaatcggtag gtaacggctagg actctggccc aatccagctcc atcaatcggct ggtcctcctcac  
1200 catcagcagcag tctatcggatcc ccttggccag cacttggccga tctcctcctgac gaccacaggcc  
1260 agtcaagctag aacgctcttctg ggttaatcag gaaactgtctc gcccctcctg ccaatcggaccg  
1320 gactcggcagc cgaagcggggt agatatacaca cctcgtctcctg ccttctcctg tgaacggccacag  
1380 tctcatgagga taacctctcac ccggtctgcca gagggtcgggga tcaaccactc gcaagctccc  
1440 gctagctgctc tctccagctg caaacaccctg tctgatccgca tcaacagctc caacgctgac  
1500 atcaaccatcg gccaccaccct gccagctcaac agaacgctgag tcaagctctc gcyccgacatg  
1560 cgtccaccaccg gttgatctatc ccaaccaggt gttcctcctcctg gttctcggctg gttctcggagca tcaacgctcctg  
1620 atgctatcccg gctcatgcttaa agaaatcctatg gaagtcaagac tgcctctctc tgcctgtcttc  
1680 gctcggctcaatc accatctccc gctgggtatagt cctgcccagctc agtctcctgct tccacaacaac  
1740 ggtgtatcagc acactctccc cggccaataaac atcaacgctcctg acactcggctc caatcggcgt  
1800 atcagcccctcc tgatctctcca tcaatctcctg atlatctgacccc cacactctgc cgtcaatgagt  
1860 gacccgctcag aaacgctcagca cgtatacctcctg gcccctcggccaa ccccaccacag atgtcccctgca  
1920 tctcagctag tctcctcctcctg gctcgcacag gaaatgcaag agcggacgaa gccaccactgg  
1980 ggtcatatata cactcctcctc gctcaaaagc ggtctcagtt taaagctgctc ccaatctacg  
2040 gttatcaaac aaggtatgcca tctggcaggt ctagtcaaat atcaaaatcg  
2100 taaaatcag tgcgggctcac tgcctctctg catcagactc tctcctcctg tctcctcctg  
2160 atcaaatcaaa tcaaatctg gttctgaggg agaaatcaac gactctggaac gactcggagac  
2220 ccatggacatc tccggaaagaa gttctctgag agtctcctatgga atgtctcctca atgtcctgctca  
2280 gacctcggaaa gttcctcgaac aatcctcaca aagagctccc gaaatcaat aatcaatctag  
2340 gtaagggcgg tccagcggga agccctcaaac caaaagctt tgbatgactg gaaaagagt  
2399 tctcatctc gattgagat gaagccgagga cgtcctcctcctg ggtatcctgat tctgtatcaaa

- cont Inued

-continued-

caatccaaga gtttcgagaa tcgaaccaaa tcccaaaa gaagtccgaa  
 660  
 aatakataat aattagata agggcgatc aggcgagaa cctaaa aagttaga  
 720  
 tgaagttaga aagagttag ataatgata tgaagttag aagcagagat cggtcgagaa  
 780  
 ttcgattcgt tactaa  
 796

<210> SEQ ID NO 17  
 <211> LENGTH: 1649  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 DNA  
 <400> SEQUENCE: 17

atggtcctag tctttaaag taagttaaat attaatgagt tactcgatct gtcaaaagct  
 60  
 gagaacttc tccctcgat gtcacgctt gtaaaagttag tactagtttc aaagttagat  
 120  
 aagattattg tccatgaaaa tcaagggtc aggcacagca catcaaaagag atcgtctgatg  
 180  
 gttatcgtgt gacggtcgcga gaaacttaca tgaaccgagag tgatccgagag cgtcgggtcgt  
 240  
 agttcaacgt tctctccgag caatggtccg caatggtccg aatatatccc gttgcaccctg  
 300  
 tccggttctgt tggcaatcat ccaatcacc agccttgggt ggttcttctg accgctctatc  
 360  
 agtctcttaa tccgtctgta gttcgtcttgc tgaagtcttccc cgtctgactgc cctctccgctg  
 420  
 tactagttctt cggcttctgt gcccgtctcg aaaccaatgc ctaaaagagag gttaaaagcgtg  
 480  
 acagcagcag tttcatcaat caccacgattg ccatggtctat cttgcaccatc gaggcatcctc  
 540  
 tcagcgttaag gttcaatgcga ggtcaagtag gattcgtccc caatcccagtc cattaatgctg  
 600  
 tggttctgtga ccaatcagcaat gttatcagat ccttctgcac gccaagtccc atcttcatgaa  
 660  
 cagtaaaagta gaaaggttctg tggttcaatca ggaactggtc gccaactcagtc gcccctcactc  
 720  
 gccaactgac ggtatgaccgag tcagaaagcggg tactaatcac acctctgtctg gcttcttggct  
 780  
 gttcatcagag atcaacttca cccgggttgc agaggttgcgt atccaaccatc  
 840  
 tgaagaatcc ccgtctagtt tctgaccagtt gttctcagag gttctcggcgt gttgtagagagc  
 900  
 tcaacgctgga catcaccatc tgcacagttca acgacgtcga gttcagagcgt gttcagagc  
 960  
 tgcgtcagacat gctcaccaca gttgatatacgt tccaaccagag tgcctcggcgt gttgtagagagc  
 1020  
 atcaacgctg gttggtatccc ggtcatagttaa aagaaatcat ggaagttaaga cttgcttcttc  
 1080  
 tctgcctgtt tctcgtgtat caccatctcc ggtcgggtatg tctcgtccagtt cagtctcgtg  
 1140  
 tccaacaaaa cgttgatatacgt tactatcttc cctgcacatcaa catagggcgt gacatcgtgct  
 1200  
 tcaaatgtcgt tactagcctcc cttgagctcc atccacttccc gatatactgaa cccaacttctg  
 1260  
 ccgttaatgag tgcaccctgcat gaaacgcaag acccgcatatc tacgcaatcctg  
 1320  
 ataaagatctt cgtcctgtatc ccaagacgttca ccaagacgttgc ccaagcaatc  
 1380  
 aactgtatcgt taaaactccc tggcacacagca atctgcccggc tcttcttctgta accgctcttccc  
 1440  
 caccacaagct gatcaatccc acagtcttccg cgtaccgcaaa gttctcgaacc aatcctcaaa  
 1500  
 aagaggtccc gaaaaataat aataatttag gtaaggtcgtg tccagcgtgga aggcctcaaaa  
 1560  
 caaaaagttc tgcattgagtt gaaaagagtt tctgatatctt gatctgaaagt gaaagcctgagaa  
 1620  
 cgtcgggtcgt ggtatctgat tctgattaa

<210> SEQ ID NO 18  
 <211> LENGTH: 3351  
 <212> TYPE: DNA

atgcatatca cacaaacagc cacatcgtcc gcttctgttg agaacgtccg aggtlaacaat 60  
accctgtctca acgatctctgc aaagcggcgt ctatatgaca cagcgtctga tgaattaat 120  
gctagggacc gcaaggcctaa agtcaattt tccaagtag taaagcgaaga acagacgtt 180  
atgtgcaacc aagatctcca attaacatct acaaacacga gaatgtctgtg 240  
cactcccttg caggcggctc ccatcctaa gaattggaat atctgactgat gcaaatctcc 300  
taccgatcat tgaatcatga taccgaggt aatttctgcat ctcatctctc caaaggcga 360  
gcatagctc accgtctgat gccgaatctg gatgtccgag acatcatctgag gcaagagggc 420  
caaaagaca gatattcaat atacccttct aggtctcagaa ggggtccgaaca acatgtccca 480  
aacctccaaa aggaagcttt cgaacgatba cgtgaaahac gctgaaahac caaacgaaat agtctgtccac 540  
gatatacttcc aaacgtctgtag gcatctcca gaatgtcca gaaatgtcca cgggaaaggt gtabgtctat 600  
gcttctcagat gctatcatagaa tacaaccctgag gcaaccctgag gcaaggtctcag gctgaaagaa 660  
aatgtacaag tatgttcatg ccgtcttcca ctttccagaa atttcacctc cgaagatcca 720  
cagttccatc ctagaagatba caatgctagt tccaagagag atgtgaaagag gttgacttct 780  
tcccttctcag ctagaagatba tcttaatat agtcaatgat atctcaatat atctcaatat 840  
gttcttgaaca cttaaccctc agcctctaat agagaaaggt acatcgaagaa gttcttcaagt 900  
actagagttc ataacctctgt tgtlaaat tctgaaatat tctgaaatat atctgaaatat 960  
ggtgtatagc atcaaggtgt agtatagtag cagtcttaca aggtctatga agaacagcagt 1020  
cactcaaaa agtctctctg gatgtctgca atctgaaagaa tctctgctaga gtabctctca 1080  
tcaattaat actgttctcc aaaaatgagg gatcatgtag tagttccatc atttgaaatca 1140  
tctctctcagat ctagaataaag gacagcgaaa gtagtctctat atctcctcaat 1200  
acagtcttca atcaatctcag tatcaatctc caaagctca caaagctctca 1260  
tcttctcgtccg aatcaatctcag tctcagagtag atcaatcaacg ggttactaacg taggtctgag 1320  
ttaggtgtatca atcaatctcag tttcaatctcag atcaatctcag atcaatctcag 1380  
cttctcgtccg tgaataaagaa tcttcttgaat agcaaggttctg cacttctgagc aaaaacctctc 1440  
tcaacaacaatctg tctgaggtatga gatcttctcca gatcttctcca gatcttctcca atctcctcaaa 1500  
gaaatgtatca taacaacaaac accgatctcaaat atctcaacaa atctcaacaa gatcaggtctg 1560  
ccgtaatctc atgtatctcag ttaggtctctg atgtatctcag atgtatctcag atgtatctcag 1620  
atcagcgtctcag tagaaatctag gaaaaatctag gaaagaaatca caatgcaatca 1680  
tccctcagaaatctg atcaatctca aaaaatctcaaat aaaaatctcaaat aaaaatctcaaat 1740  
tctcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 1800  
aaagcgtatcaatctcag tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 1860  
tctcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 1920  
gaaacccatcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 1980  
gacttctctcag atcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 2040  
cctcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 2100  
cctcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 2160  
gtctctcagaaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 2220  
aaagcgtatcaatctcag atcaatctcaaat tcaatctcaaat tcaatctcaaat tcaatctcaaat 2280

<400 > SEQUENCE: 18

<213 > ORGANISM: Tomato mosaic virus

- cont Inued

2340 catgcatggg gggctgtgta gactcatgag aggaatatc acgtcgcacat actgagacac  
2400 gatgaaatctg gcatatctac tggcagatbac tggcagaggg tgcctgtgag tctctgagctg  
2460 gtagtatatct ctgatatactg taactcaggg actctgagaa gattctcaca agtatggagaa  
2520 ccacacagta gtcacagcaaa gctggttttg gtagatggcg tccagggctg cgggaaagaca  
2580 aagaaatct tctcagagat taattctgaa gaagatctaa tctctgtccc tggctcgtcaa  
2640 gctgctggagat tgcacagagat aagagctaat gctcgggca taatagctgct tacaaagatc  
2700 aatgctgctg ca cgtcagatc attttctgag aatctacggga aaggggcagc cgtcagatc  
2760 aaagatctg tcatagacga agttctgag ctgcatctg gctgctgagaa tttttctggt  
2820 gaaatgctctc tgtgctgat tgcatactgt tatggagaca ccaacagat tctgcatatc  
2880 aaacagagtaa ctggttttcc gtaaccctgca cactctgcaaa aatctgaggt cgaacagatc  
2940 gaacacagaa gaactatctc tctgtgtccg gctgtgactca cacactccc aatcacaagg  
3000 tatgaaagac agctaatctg cacgtctct gaaagaatct cagtttccc agtaaatggt  
3060 agtgggggctg cgtctatcaaa tccctgtgccc aagcccgtca agggaaatct tctgactctc  
3120 acaacagctg acaaggggccc ccttctctca aggggctatcag cagatctcca tactgtacaat  
3180 gaggtaacaag gtagaactaa tgcagacgtat tctgtagttc gactcaaccct tactgctgta  
3240 tctatctcag caagagacag tccgctatgt ctggtctctcgt tgcacaagaca cacaaatccc  
3300 ctaaaatctc acaaccgtgt gatgactctc tagttatgaa tcaatagagaa tctagaaaggg  
3351 gttagtagt acttatagaa catgtacaaa gtagatgag gtagatgag gtagatgag

<210> SEQ ID NO 19  
<211> LENGTH: 4851  
<212> TYPE: DNA  
<213> ORGANISM: Tomato mosaic virus  
<400> SEQUENCE: 19

60 atgcatcaaca caacaacgc cactctgccc gctctgcttg agaacccgtccg aggtacaacaat  
120 accctgctca acgatctctg aagcggcgt ctatatgaca cagcgtctcga tgaatttaat  
180 gctaaaggccc gcaaggccc aatcaatctc tccaaagtag taagcggtaga acaagacgctc  
240 atctgcaaca aagcccatac attaacatctc acaacaaccga gaaatgctgctg  
300 catctcccctg caggcctcat cctgactat gaattctgaa atctctgatat gcaaatctccc  
360 tactgtagtca tgaactatgaa tactcggtagt aatctttctgcat cctcactctgtr caaaggctgga  
420 gcatatctc actgctctg atctgctat gctgactctg gatgctccgg acatcaatctgctg  
480 caaaagaca gtagtgaact ataacctctc aggtctcggaga ggggccaaca acatgctccc  
540 aactcccaca agaaagctct cgaacagatc gctgtgaaatc gctgtgaaatc caaaacgaaat agtctgctccc  
600 gatcatctcc aaacctgtag gcatctcaca gaatgctaca cgggtagagat gtagtctcat  
660 gctctctgcat gtagtatcaacgaa tacaaccctgca gacagatctcgg gctcgtgaaag  
720 aatgtagatca tagtctatctg cgtctctccc cgtctctccc atctctcccga atctcaatctc  
780 caagctcaacc tccgacagat caatctgcatg tcccacaagg atctgagagag gctgactctct  
840 tccctctgcat ctagtagaac tcttaaatat agtctcaatgat atctcaatgat tctctaaagtat  
900 gttcttctg caaaaaccctacac agcccatac agtagagagttc acaatctctctc atctgtagta  
960 actaagagttca ataacctccctg tctgtagaattc tctagaaatctc atctctctctc atctgtagta  
1020 ggtgtagcagcc atcaaggtgtg ataatagtag agtagtagag cagttctcaaa aggtctctctg  
1080 caactcaaaa agaacctctg cgtctctctg gatgtagcaca agtctgtagcaca agtctctctc



780 gtagtctgaa aagagttctga taattctgatt gaaagatgaaag cggagaccgtc ggtccgctgatt  
720 aataataataa atcttagtga aagggcgtca ggcgggaaag ctaaaccaaa aagttcttctgatt  
660 gttcccaatgt cgtttagacc cgcacaagtct cgaacccaaat cctcaaaaag aggtcccgaa  
600 gtagaacgatt gtagaccaccat ggaaccctcg gaaagagttg ttgattgagttt catgtaga  
540 tctgtctgtat tcttctatata aataataataa aattctggtt ttgaggtgagaa agtcaacgatt  
480 gtaataataa aataatgtaaa aatgagttcgg ggcctacatc cttctgtccat agaatcttctg  
420 gttggtcccaa artcaacggtat tacaacaag gttcagaaag gttcagaaatg gcaagttctca  
360 gaaagaaacca caactgggtc atactacaact gctcgtctgta aaaaaggttctc tcaagttcaaa  
300 ccagatcaat gcccgtgtgtt ttgtagttctc ttgcattggtt accaagagatt ggaagagcgg  
240 cttatagaa gttggttatt ttgcttagtc gttccttctg ttgcccgttctg gttggaattca  
180 aagattatggt tccatgaaaa ttgatcattg tctgaaagtta atcctctcaaa agttctgtaaaa  
120 gagaaactc tcccgtctgat gttcaacgct gtaaaagttg ttatggttcc aaggttctgatt  
60 atggtcctag ttgttaaaag taagttcaat attaattgatt ttactgattct gttcaagttct

<210> SEQ ID NO 20  
<211> LENGTH: 795  
<212> TYPE: DNA  
<213> ORGANISM: Tomato mosaic virus  
<400> SEQUENCE: 20

4851 tcaagataaag tttctgttag aagttctatt cttttagtgc ttagttctga a  
4800 gtagttctata aaaaaccgccc accctggttctg tttgtttata agagtttattg taagtattctg  
4740 gatgttctctg agtccgttata caattcggcgg tatttacaaca caattgagcga cgtctgttggg  
4680 cttgtgtctca aacacatcaa gtagttggat cacttggagg agttccaggag atcccccctgt  
4620 gttattctac accgatagagg ttgcattaga taccatagacc cttgaaagt gatttctgaaa  
4560 atgtggatatt ttgagggcaa actgttcaag aagcaatattg gttactcctg cgggaggtacc  
4500 agttctgtt actttcccaa ggttctgttag ttaccaccgata tacaacaagc tgcataatcca  
4440 ttcttagcatt caatgctccc gatgtaaaa ttgatcaaaag gaggcctcctg cggagattgac  
4380 taccagagaga agagttgttga ttgtacaact ttatctcggtt ataccggtcat catctgctcgg  
4320 caaggtgcatg aaaaaaacca cctgaaagt taccactgctg gttacaaaaac gttgtttatgg  
4260 gtttagtaccg aatcctgttag gtagcttgggt cttgagagatt tcttggcagg agttgtggaaa  
4200 gacgtactctg agttgttagt ttccagatatt gataagttcc aaaaagattc tcatgttctgt  
4140 agaaagacac ccgctccagat cgaagatttc ttccggagatc tagaacgtcca tgtcccacgtg  
4080 agttgagctca caaggtcaat acttgaactt atctgacctcaa gcaattctt gttccttaccg  
4020 gcttctgcaaa cgtattgttga tcaatcaaaag aaatcaacgg caatatttgg tccctcttctc  
3960 catattgattt aagccgcaacc gaaagcaaaa ctggtatctgt caattcagac agaatatcca  
3900 acaattggtc agttcggccga tttttagattt gttggtatctc cagccggttga tcaagtacagg  
3840 aattttctac tttttagtag agagttctcc aataggttga tagcaaaagca agaaacaagtcc  
3780 ttagtgttag ataattttt ttatagttat ttaccttaagg aaaaagaaa accaaacaaa  
3720 aaaaagaatt tcaattcacc agagttctcc gtagttagttg atattgaaa taccctgcatc  
3660 cgaaccgctgg cagaaatgccc ttccagatc ggaactgttgg aaatctcatt tgcgtattat  
3600 gatattgtcca agtctgttagc gattgtcaaac gattgtcaaac caacttcaat accgtattga  
3540 tagaacctg ttaaccatga atcttctctga atgtcaaaag ttgcattata

-continued-

795

ctgattctgt attaa

<210> SEQ ID NO 21

<211> LENGTH: 1336

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 21

60 tggattttgc cctattttcc cagacatctg ttatcactta acccattaca agcccgctgc

120 cgcagatatt cccgtggcga gctataaac acccgcacat agcggatgcca tccgctataa

180 tgcctcgaacg cctctgcag gtcctcttgc gccgttaca ccgctcgtttc gggcactgata

240 cctgattgagt cagcctttat cgttttccag aagctcctctg ctattcctgt acctcctggg

300 cctccgcaccg ccgtgttctt ccgtttccagt ccaaacatcc gggcgcagctg gcgtgtttca

360 ttaccgccgtt agcattgaa accatccctc agccacccca ctggagagcga cggaaagtccg

420 ttgcctgaagc ggtgttccaa ccgtccccaagc atgaccgttcc gtactgtttc acctgttgaa

480 cccgcctgttag tgaaccgcc a gtcagatgccc tcaagatcac agcagttccag cgcgaaacgtg

540 acaagcagtc tctctctcgt atcaacagcag aactcgcgaa ccgtcagagca ccatcctgcctg

600 tctctttctt tcaaccgcc a tctcgtctgta tctgtccctgt tcaagcaggt

660 tttcctgcctc aacaaaccgc attcctggcgc atgtatccggc aaacacgctt gccatctgatc

720 gcaaggcattc cactcaagtc tgcctcctgt cgaagcagcc gccattacc ccgaaatacca

780 taagtgtgca gctctctcgt aacatgtgtg atacgtagga gcaaccatccgt atcactcagtg

840 tgaagcactgc ggcctggccac caccaccagt tccgttccga tggagatcac gtcgcaaccgc

900 gcaagcagca cccggagaca accgctgact aagcttacc cccatccccc ggcataaagg

960 gcctgtgctgc taccaccatt tctcgcctgc cactattcaaac ggtctctctg aggagtctcat

1020 tttcctcctg tttctctgc acccgtgcgc gtaggttctt aatcctgctc atgtgcctcag

1080 caagttccga ggcaggaaca accctgttcc cggcggcga accagttaga cttcctctcc

1140 ggtattgtct acccagagaa aataactgac tggctcgtca accatgctgc cgggcaacga

1200 gggagagacc cactccctgt tcaagctct gctgaaacaat tgcgactctt tccctgtgtg

1260 taagcctgct gcgttctcc ggcctcaaga cactcaatcat ctgttctcca atgaccctagtc

1320 taaaaactag tattaagacc atccattat taagtgtatat tggltgtctg gagatctcagg

gggcccagct agtga

1336

<210> SEQ ID NO 22

<211> LENGTH: 1258

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 22

60 tggattttgc cctattttcc cagacatctg ttatcactta acccattaca agcccgctgc

120 cgcagatatt cccgtggcga gctataaac acccgcacat agcggatgcca tccgctataa

180 tgcctcgaacg cctctgcag gtcctcttgc gccgttaca ccgctcgtttc gggcactgata

240 cctgattgagt cagcctttat cgttttccag aagctcctctg ctattcctgt acctcctggg

300 cctccgcaccg ccgtgttctt ccgtttccagt ccaaacatcc gggcgcagctg gcgtgtttca

360 ttaccgccgtt agcattgaa accatccctc agccacccca ctggagagcga cggaaagtccg

420 ttgcctgaagc ggtgttccaa ccgtccccaagc atgaccgttcc gtactgtttc acctgttgaa

480 cccgcctgttag tgaaccgcc a gtcagatgccc tcaagatcac agcagttccag cgcgaaacgtg

-continued-

540 aacagcagtc tctctccggt atccacagcag aacctcgaacc cgtcagagacc ccatcgcctga  
 600 tctgcttctct tccacggccac tctgctctga tctgctccggt tccagctgagc  
 660 tctcgtctcaaa gcaaacagcgc atctctggcgc atgctccggt aacaacgctt ggcctctgatc  
 720 gcaaggcctaac cctcaagcttc tgcctcgtctg cgaagcagcgc cccatcaaccg accgatcaacc  
 780 tcaagcttgcca gctctccgat aacatggctgt atcacggagaa gcaacatccgt atcctcagtg  
 840 tgaagcactgc ggcggccatc cctccagctca tccgctcgtc tgaagcaatgc gctgcaaacctg  
 900 gcaagcagcaacc accggctgact aagctctaac cccatccccc ggcacaatcaagg  
 960 ggcgctgctgc taccacctct tctgcccgc ccatctcaac ggcctctcttg aggagctctcat  
 1020 tctccatcgtt tctctcgcgc agcagcgcgt ggaagctctt atctcgtcc atgctcgcag  
 1080 caagctctcag ggcaggaaca accctgctctc cggcggcgcac agcagctaacg actccctcct  
 1140 gctatctgctc agccagagaa ataacctggc tggctgctac accatgcttg ccggccaacga  
 1200 gggagcagcgt cactcccgcgt tcaagctctc gctgaaacat tgcgactctt tccctgctgag  
 1258 taagcgcgtc gcttctcc cggccctaacg cactcaatcat cgtctctcca atgctgag

<210 > SEQ ID NO 23  
 <211 > LENGTH: 1333  
 <212 > TYPE: DNA  
 <213 > ORGANISM: Escherichia coli  
 <400 > SEQUENCE: 23

60 taagcttca ggcacagcac atcaagagaa tccctgatg tatcctgctg agcctcgcag  
 120 aacatcaatc tgaagcaggt gatccgggacgc gtcgggtctgaa gttcaacgctc tgcctcgcgc  
 180 agtggcgcga aatctccccc tgcacctcgc ggaaggtgat cccggtctcgt ggcacatcaccc  
 240 caacatcaacc agctctgggctg gcttctctca cgcgctctaat gctcctctaat cgcctgtaag  
 300 tgcctcctgct agcttctccc gctgactgccc tctcctcgcgt acaagctctct cggctctctg  
 360 cccgcctcctg aaccaatgc taaagagaggg tcaagcgcga cagcagcagct tctcctcaatc  
 420 aaccaatgatc catgctctcatc tgcctccagctc agctatcctc cagctcgaagg gtaatgctgag  
 480 gttcaagctgg agcttctggcc accctccagctc atctaaatgctg ggtctcgtgac ccatcgcagc  
 540 tctatcagatc cttctgcccag cagctccagc tctctcaatgc gaccacaagcc agtcaaaagtga  
 600 aacggcttctg ggttaaatcaag gaaactgctcgc cccctcaatgc ccaatgacccg gatgctcgcagc  
 660 cgtatctcaac agatctcaacc cctcgtctctg cctctgctctg tcaacagctctc tgcctcgcgc  
 720 gcccaccccag gctccagctcaat aagctctaac cccatccccc ggcacaatcaagg  
 780 gcaagcttgcca gctctccgat aacatggctgt atcacggagaa gcaacatccgt atcctcagtg  
 840 tgaagcactgc ggcggccatc cctccagctca tccgctcgtc tgaagcaatgc gctgcaaacctg  
 900 gcaagcagcaacc accggctgact aagctctaac cccatccccc ggcacaatcaagg  
 960 ggcgctgctgc taccacctct tctgcccgc ccatctcaac ggcctctcttg aggagctctcat  
 1020 tctccatcgtt tctctcgcgc agcagcgcgt ggaagctctt atctcgtcc atgctcgcag  
 1080 caagctctcag ggcaggaaca accctgctctc cggcggcgcac agcagctaacg actccctcct  
 1140 gctatctgctc agccagagaa ataacctggc tggctgctac accatgcttg ccggccaacga  
 1200 gggagcagcgt cactcccgcgt tcaagctctc gctgaaacat tgcgactctt tccctgctgag  
 1258 taagcgcgtc gcttctcc cggccctaacg cactcaatcat cgtctctcca atgctgag

```

<210> SEQ ID NO 25
<211> LENGTH: 1338
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli
<400> SEQUENCE: 25
60   taaggttca ggcacagcac atcaaaagaa tgcgtgatgg tatcgggtgag agcgtccgcag
120   aacattacat tgcacagcaggt gatccgggacgc gtcggggtcga gtttaccggtc tgcctccgcgc
180   agtggcgcgc aatattcccg tgcaccctggc ggaaccgggtat ccgggtctcgtc ggcacaatcctc
240   cacatccacca cgccttggggtg gtttcttgcga cgcgtctatca gctccttcaat cgcctgttaag
300   tgcgtctcgtc gaggttccccc gttgatcctgcc tcttcgcctgtc acagttctctc cggccttctgtc
360   ccgcctctcga aaccaatgcc taagagagaggg taaagaccgcga cagcagcagctc tccatcctcctc
420   accacagatgc catgttctcacc tgcctccagtc agcctcctctc cagcgttaaagg gttaatgctgag
480   gtaaggtaagg agttggcccc aatccagttcc attaatgctgt ggtcgttgcac ccatcagccacc
540   ttatcgaatc ccttgcaccag caagttccgcga tcttctcatgac gccccaagacc agttaaagttag

```

```

<210> SEQ ID NO 24
<211> LENGTH: 1333
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli
<400> SEQUENCE: 24
60   taaggttca ggcacagcac atcaaaagaa tgcgtgatgg tatcgggtgag agcgtccgcag
120   aacattacat tgcacagcaggt gatccgggacgc gtcggggtcga gtttaccggtc tgcctccgcgc
180   agtggcgcgc aatattcccg tgcaccctggc ggaaccgggtat ccgggtctcgtc ggcacaatcctc
240   cacatccacca cgccttggggtg gtttcttgcga cgcgtctatca gctccttcaat cgcctgttaag
300   tgcgtctcgtc gaggttccccc gttgatcctgcc tcttcgcctgtc acagttctctc cggccttctgtc
360   ccgcctctcga aaccaatgcc taagagagaggg taaagaccgcga cagcagcagctc tccatcctcctc
420   accacagatgc catgttctcacc tgcctccagtc agcctcctctc cagcgttaaagg gttaatgctgag
480   gtaaggtaagg agttggcccc aatccagttcc attaatgctgt ggtcgttgcac ccatcagccacc
540   ttatcgaatc ccttgcaccag caagttccgcga tcttctcatgac gccccaagacc agttaaagttag
600   aacgggttctt gtttaaatcag gtaactgttcc ccttccactgc ccaacttgaccg gattgcggaccg
660   cgaagcggggtt agtatctacaca cttcttgcctgg ctttcttgcctg tgcaccgcacag tccatagtagag
720   taacccttccac ccgggttgcacca gagggttgcggaa tccaccactc gccaaagctcc gctagtgcctc
780   tgtccagttc gaaaccaccctc tgcataccgcga tccacgcagtt ccaaccgctgac atccaccactg
840   gccaccaccact gccactgcacc accagttctgg ttaacagttctc gtcgcgacatg cgtccaccaccg
900   gttgatctatc gccaccaccagg gttccggcgtg gttgtagagagca tcaaccgctcgtc atggtatcccg
960   gcatatgtcaa agaaatcctag gaagttaagaa tgcctcttctc tgcctgttctc gtcggtctaac
1020   accattccccc gtcgggtaatatc ctgcaccagttc agttcctgttctc tcaacaacaaca agtgtatctagtt
1080   accatttctcc ccggccaatcac ataccggcgtc accatcggctc caaatgggctc caaatgggctc
1140   tgatgcttccca tccacttccctg atctatgaccc caccactctgc cgtaatgagtt gaccgcgcactg
1200   aaacgcagcaga cgtatagcctgc gcttccggta taagagactcc gccgtctgtatac
1260   cagatctgttc ccgcataatc agaatatctc gcatcgggcga actgtactgtc aaacctgcctc
1320   ggcacacagcaa ttgcctccggct tctcttgttaac gtcgtcttccc accaacgctgc atccaatctcca
1333   cagtttctgc gat

```

600 aacggctcttctg ggtctcaatcag gaaactcctctcgc cccactccactg ccaactgaaaccg gactgaccgacg  
660 cgaagcgggggt agtatataca caactctctctg cctctctctgctg tgaaccgcaacag tccatagagaa  
720 taacctctccac ccgggtctgcca gaggctgaccga tccaccactt gcaaaagtcccc gctagctgctc  
780 tctccactgctg caaccaccactg tcaaccgactt caaccgactt caaccgactt caaccgactt  
840 gcccaccaccct gccagttcaac agaacgctggt tcaacagttctt gctcagaccact gctcaaccaccg  
900 gtttatatactgt ccaaccaccagt gttctcggctggt gttgtagagagca ttactgctctggt atggtatctccg  
960 gcatatatacaa agaaatcactg gaaagtcaagac tgcctctctcttctc tgcctgctctctc gctcggctatca  
1020 accactctcccg gctgggtatagtt ctagccaccagt cagttctctgtt tcaaccacaac gttgtagatca  
1080 accactctctcc ccggtcaatatac ataacgctggtt accactcctgctt caaatggtcgtt atagcctgccc  
1140 ttagctctgctc caactctctc atctatctctg atctatatac ccaacccttctg cgtatagttggt gaccctgctc  
1200 aaacgcaagca cgtatccctg gcttctcggta taaagacttc gctcctgatac  
1260 cagatctctgct ccgcatataat acgataatct gcatcctgctga actgatactt aaaaactgctc  
1320 ggtcaaccagca tctgcccctt tctctctgtaac gctgctctcc accaacgctg atcaatctca  
1338 cagttctctgct gactgtagg

<210> SEQ ID NO 26  
<211> LENGTH: 1333  
<212> TYPE: DNA  
<213> ORGANISM: Escherichia coli  
<400> SEQUENCE: 26

60 taaccgctca ggtcaagagaa tccctgtagt taccctgtagt atcaaaagagaa tccctgtagt agctcctgctc  
120 aacatatacat tgaaccgcaagt gaccctgagc gttccgagagc gttccgagagc gttccgagagc tgcctccctg  
180 agtggtagcagca aatatactccg tgaacccttgc ggaaccggttat ccggttctgtt ggcacaataactc  
240 cacatataccca ccgctctgggtg gttctctctca cctgctctcaat gctcctctcaat cctcctctcaat  
300 tctgctctctgct gactgctctctc gttcagctctgccc gttcagctctgccc gttcagctctgccc gttcagctctgccc  
360 cccgctctctgca aaccactgccc taaagtagaggg tcaaaagcctga cagcagcaggt tccatcaaatc  
420 accaactgctct cactgctctca tgcctccactc agtccactcctc agtccactcctc agtccactcctc  
480 gttcaggttagg agttctgccc aatccactcc atctatactggt ggtcctgctgca cactcagaccg  
540 tcatctcaatcc cttctgcccac caagttccgga cctctcaatgaa gaccacaagcc agtcaaaagttag  
600 aacggttctctt gttggttcaatcag gaaactgctctc cccactccactg ccaactgaaaccg gactgaccgacg  
660 aacggtatataa agaaatcactg gaaagtcaagac tgcctctctcttctc tgcctgctctctc gctcggctatca  
720 gtttatatactgt ccaaccaccagt gttctcggctggt gttgtagagagca ttactgctctggt atggtatctccg  
780 tctccactgctg caaccaccactg tcaaccgactt caaccgactt caaccgactt caaccgactt  
840 gcccaccaccct gccagttcaac agaacgctggt tcaacagttctt gctcagaccact gctcaaccaccg  
900 gtttatatactgt ccaaccaccagt gttctcggctggt gttgtagagagca ttactgctctggt atggtatctccg  
960 gcatatatacaa agaaatcactg gaaagtcaagac tgcctctctcttctc tgcctgctctctc gctcggctatca  
1020 accactctcccg gctgggtatagtt ctagccaccagt cagttctctgtt tcaaccacaac gttgtagatca  
1080 accactctctcc ccggtcaatatac ataacgctggtt accactcctgctt caaatggtcgtt atagcctgccc  
1140 ttagctctgctc caactctctc atctatctctg atctatatac ccaacccttctg cgtatagttggt gaccctgctc  
1200 aaacgcaagca cgtatccctg gcttctcggta taaagacttc gctcctgatac  
1260 cagatctctgct ccgcatataat acgataatct gcatcctgctga actgatactt aaaaactgctc  
1320 ggtcaaccagca tctgcccctt tctctctgtaac gctgctctcc accaacgctg atcaatctca  
1338 cagttctctgct gactgtagg

1333

cgatttcgc gat

<210> SEQ ID NO 27  
<211> LENGTH: 1333  
<212> TYPE: DNA  
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 27

taacggttca ggcacagcac atcaagagaa tcgctgatgg tacgctcgcag

120 aacattacac tgaacgaggt gatcggagac gtcggggtcga gtttacgctg

180 agtggcgcga aatattcccg tgacccttgc ggaacgggtat cgggttctgc

240 cacattcacca cgtcttggtg gtttttttca cgcgttatca gctcttcaat

300 tgcgcttctc aggtttcccc gttgacctgc tcttcgctctt acagttcttt

360 cccgcttcga aaccatggc taagagaggg tcaagccga cagcagcagt

420 aaccagatgc catgttcatc tgcaccagtc agcatctctc cagcgttaagg

480 gtaacggttag agtttggccc aatccagttc attaaattgctt ggtcgttgcac

540 tcatcgatc ctttgcaccg caagtccgca tcttcaatgac gaccaaagcc

600 aacggtttgt gtttaattcag gaaactgttc ccttccactg ccaatgaccg

660 cgaagcgggtt agatattaca ccttgttctt ctttcttgcg tgaacgcagag

720 taaccattca cccgttgcga gaggttgcga tccaccactt gcaagttccc

780 tgtccagtty caaccacccty ttgatccgca tcaacgagtt atcaaccacty

840 gccaaccacc tccagttcaac agaacgcttctt tacaagttctt cgtcaccacc

900 gttgatcatc ccaaccaccgt gttcggcgtt gttgtagagca ttacgctcgg

960 gcatagttaa agaaatcaty gaagttaagac tgccttttctt tgccttttct

1020 aaccattcccg gtcgggtatagt ctcgaccagttc agttctgttctt tcaacc

1080 acaatttccc cggcaatbaac ataacggcgtt caaatggcgtt atagccgccc

1140 ttgatgcttca tcaactcccty attatgaaac caaaccttctc cgtatgagtt

1200 aaacgacagca cgtatagcty gcttgcaccaa ccttcggtta taaagacttc

1260 cagacgcttc ccgcattatc agaatatctt gcatcggcga actgatctgt

1320 ggcacagcaa ttgcccggct ttcttgtaac ggccttccc accaacgcty

1333 cgatttcgc gat

<400> SEQUENCE: 28

<210> SEQ ID NO 28  
<211> LENGTH: 1333  
<212> TYPE: DNA  
<213> ORGANISM: Escherichia coli

taacggttca ggcacagcac atcaagagaa tcgctgatgg tacgctcgcag  
60 aacattacac tgaacgaggt gatcggagac gttcggtcga gtttacgctg  
120 aacattacac tgaacgaggt gatcggagac gttcggtcga gtttacgctg  
180 agtggcgcga aatattcccg tgacccttgc ggaacgggtat cgggttctgc  
240 cacattcacca cgtcttggtg gtttttttca cgcgttatca gctcttcaat  
300 tgcgcttctc aggtttcccc gttgacctgc tcttcgctctt acagttcttt  
360 cccgcttcga aaccatggc taagagaggg tcaagccga cagcagcagt  
420 aaccagatgc catgttcatc tgcaccagtc agcatctctc cagcgttaagg  
480 gtaacggttag agtttggccc aatccagttc attaaattgctt ggtcgttgcac

-continued-

540 tcatcagatc cttcggccag caagtccgca tctccatgac gcccaagcc agttaaagtag  
600 aacggtrtctg gtrtaatcag gaactcactg cccctcactg ccaactgaccg gatgcccagc  
660 cgaagcgggtr agtatcaca ca ctctgtrctgg ctttctgctg tgcacgcccag tccatagagga  
720 taaacttca cccgttctgcca gaggtgcccga tccaccactt gccaaagctcc gctagtgcct  
780 tgtccagtgc caaccaccctg tctgaccgca tccacgcagtc caacgctgac atccaccactg  
840 gccaccaccct gccagtcaca acc agtagctggt tcaagctctt gcgcccagcatg cgtcccaccag  
900 gtrtatctctg ccaaccaccgtr gttccggcgtg gtrgtagagca tcaaccctgag atggtatccg  
960 gctatctctg agaaatcag gaaagttagac tgccttctct tgccttctct gtrcgttatcc  
1020 accattcccg gctgggtatagtr ctgcccagtrc agtrcctgtrt tcaaccacaac gtrgttatagtr  
1080 acaacttccc ccgccaatcaac ataccggcgtg accatccgctt caaatggcgtt ataagcccgc  
1140 ttagtgcctcca tcaactccctg attatgacc caacccttgc cgtaatgagtr gaccgcactg  
1200 aaacgagcaga cgtatagctg gccctgcccaca ctttccggtat caaagactcc gcgctgtatca  
1260 cagatcgttgc ccgcatatct agaatatct gcatccgagca actgtactgtr taaactgcct  
1320 ggcacagcaca ttgcccggct tctctgtaac gcgcttccc acccaagctg atcaattcca  
1333 cagtttctgc gat

<210> SEQ ID NO 29

<211> LENGTH: 600

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 29

60 ccaactctg atccgactca ccagttcaca cgtcagatcc accatctgccc accaccctgccc  
120 agtcaaacagag ccgctgtgttca cagtrccttgcg cagacatgctg tccaccagctg atbatctgtccc  
180 ccaaggtgtct ccgcccgtgtgt ttagagcattca cgtcccccgtg gatcccggcga tagttaaagg  
240 aatcactgga gtaagactgc tcttctctgc cgttctctgc gtrtaatccc atcccggcgg  
300 ggtatgtctgc ccagtrctagtr tccgtgttca cacaaggtg gatcagttca ctttcccggg  
360 caatcaatca ccgcccgtcaca tccgctcaca atgctgtata gcccccctgca tgcctccatca  
420 cttccctgat attgaccacc acccttgcgt attgagtgac cgcactcgaac cgcagcagca  
480 tagctggccc tgcaccaccct tccgttatca agactctcgg cgtgatccag agtctgcccg  
540 catatctcag aatctctgca tccgcccgaact gatccgttaaa actgcccctggc accagcaatg  
600 cccggcttcc ttgtacaagcc ctttcccacc aatcccagag ttttccgctgat

<210> SEQ ID NO 30

<211> LENGTH: 300

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 30

60 caatcaatca ccgcccgtcaca tccgctcaca atgctgtata gcccccctgca tgcctccatca  
120 cttccctgat attgaccacc acccttgcgt attgagtgac cgcactcgaac cgcagcagca  
180 tagctggccc tgcaccaccct tccgttatca agactctcgg cgtgatccag agtctgcccg  
240 catatctcag aatctctgca tccgcccgaact gatccgttaaa actgcccctggc accagcaatg  
300 cccggcttcc ttgtacaagcc ctttcccacc aatcccagag ttttccgctgat

<210> SEQ ID NO 31

<211> LENGTH: 100

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 31

tcggcaact gatcgtaaa actgctgca acagcaattg ccggcttc ttglaacgca

cttccacc aagctgata aatccacag tttcggat

<210> SEQ ID NO 32

<211> LENGTH: 1604

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 32

tcattgtttg cctccctgtc gcggttttc accgaagttc atgccaagttc agcgtttttg

120 cagcagaaaa gccgccaatc atatgccc atbgtttg aagatccctt tcttgttacc

180 gccaaagcc atbgtccc gcaatcggc aaatccatb aatbcccata cctgttccacc

240 gacagagccg ctgacagccat caagagccg gttatataca tccagccatg cacactgtata

300 cctctacc caccatgttc gttacatgta gttgacagcccg gctaaagttat ccaagccgta

360 tccgtgtgata ataatccgtt cctccgccc gcccagatc cttttcccag

420 taactctct gctgtttccc aatccgccc ttggaatata cactccgttat aacggttccag

480 gcccagccca tcaagagat ccctgtatgt atcggttgtga atcggttgtga accatcatc

540 gaccagctg atcggaccg tccgggtcga ttacagctt gctccccc gttggtcggaaa

600 atatcccgt gacccttgcg gatcgttatc cgttccgtt gcaatatacc accatccacc

660 gcttgggtt ttttttacc gcttatatac gcttataatc gcttataatc gcttataatc

720 agttcccgc ttgactgccc cctccgctta cagtccctc ggccttctc cgcctcggaa

780 accaatgccc aaagagagtt taagccgac agcagccagtt tcaatcatca ccaagatgccc

840 atgttccat gccccagttc gacatcctc agcgttaagg gttatcggag taccgttagga

900 gtttggccc atccagttcca tcaatgctg gttccgctc atccagccat taccgaaatc

960 tttgcccacc aagttccgat cttcatgacc accaaagcca gttaaagtat accggtttgtg

1020 gttatcatgg aactgttccg cctccatcgc cactgaccgc atgcccagcc gaaagctgta

1080 gatataccac tctgtcctgc ttttggctgt gacgccaagt tcaatagagat aacctccacc

1140 cgttctgcag agttgctgat tccaccatc caaagttccc gttgctccgt gttccagttgc

1200 aaccaccctt tgaatccgcat cacgctgacc aacgctgacc tccaccatcgg ccaaccaccg

1260 ccagtccaca gacgctgctt tacagttctc cgtaccatgc gttccaccacg ttgatatcctc

1320 caaccaccgt tccgctgctg ttgtagagcat tacgctgctg ttgtagatccg catagttaaa

1380 gaaatcatgg aagtacagct gcttttctt gcttttctt gcttttctt cgtgttaatac

1440 cgggtatagc ttgccaagtt gttcgtttgt caccacaacg gttgatatcga cactttccc

1500 ggcataatac tacggctgta catcggcttc aaatgctgta tagccgccc gatgctccat

1560 cactccctg tcatrtgacc accatctgcc gttatctgct gttatctgag accgcatcga aacgcaagcc

1604 gatacgtcgg cctgcccacc ccccacaaga ttctcctgat ttga

<210> SEQ ID NO 33

<211> LENGTH: 480

<212> TYPE: DNA

<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 33

ccgtggacca caatccaac aagaaacac aaaaagctt ctatgagatc tcaacttacc

60

```

210 > SEQ ID NO 34
211 > LENGTH: 1333
212 > TYPE: DNA
213 > ORGANISM: Escherichia coli
<400 > SEQUENCE: 34
60 taacggttca ggcacagcac atcaagagaa tgcgtgatgg tatcgtgtgtag agcgttcgtag
120 aacattacat tgaacgtaggt gatcggtagc gtccgggttcgaa gtttacgtagt tgcctccgccc
180 agtggcgcgaa aatattcccg tgcacccttg ggaacgggtcat ccggttctcgt ggcatacatcc
240 cacatccacca cgtcttgggtg gttttttgtca cgcgttatca gctcttcaat cgcctgtlaag
300 tgcctctgtc agtttcccgc gttgacctgc tcttcgctgtt acagttcttt cggctctgtg
360 ccgcctcgcg aaccatgccc taagagaggg tcaagccgaa cagcagcagt tccatccatcc
420 aaccagatgc catggtcatcc tgcaccagtcc agcattctctc cagcgttaaggt gttatgtagag
480 gttacggtagt agtttggccc aatccagttcc attaaatgctt ggtcgttagaac catcagccacc
540 tcatccgatac cttcgcacag caagtccgca tctccatgac gaccacaagcc agtcaagtagt
600 aacgggttctt gtttaartaag gtaactgttcg cccctccactg ccaactgacag gattgtagcag
660 cgaaggggtgt agtatccaca cctcgttcttg ctttctgtcgt tgaacgacag tccatagtaga
720 taaacctccac ccggttgcga gaggttgcgaa tccaccactt gccaagctcc gctagtgtccc
780 tgtccagttg caaccaccctg ttgatccgca tccagcagtt caacgctgac atccaccactg
840 gccaccaccct gccagttcaac agaacgctcgt tacaagttctt gcgcccacag cgtccaccagc
900 gttgatcatcgt ccaaccccaggt gttcggcgttg gttgtagagca ttacgctcgtg atggtattccc
960 gcatagttaa agaaatcatg gaagttaagac tgccttcttcc tgccttcttcc tgccttcttcc
1020 aacctcccgc gctgggtatagt ctcgcccagttc agttcgttctt tccaccacaac ggtgtatcagt
1080 acaacttccc ccgccaatbaac ataacggtcgtt acatcggcgtt caaatggcgtt atagcgcgccc
1140 tgaatgctcca tcaactccctg attatgaaac ccaacttctgc cgtaatgtagt gacccgcatcg
1200 aaacgtagca cgatcagctg gctcgtcccac ccttcggtaa taaagacttc gcgctgtatca
1260 cagatcgttc ccgcatcatc agtatcatc gcatcggcga actgtagtctt aaacctgctc
1320 ggcacagcaa tgcgcccggct ttcttgttaac ggccttccc acacaagctg atccaatccca
1333 cagtttctgc gat
ctacaactac aatggcatca accacaacag ccaactcgtc cgccttgcctt gtagaccgtcc
60 gatttttacc aacatltacc aacaacaaca aacaacaaca aacaacaaca attttatcatc
400 > SEQUENCE: 35
210 > SEQ ID NO 35
211 > LENGTH: 6384
212 > TYPE: DNA
213 > ORGANISM: Tomato mosaic virus

```

180	gagrtatcraa taactrtgtrc aacgatctrg caaagcrgcrg tctatrtacrg acagcrgtrcrg
240	atgatrttaa tgcragggac cgcagggcra acgtrcaatrt tccaagrtat gtaagcrgaag
300	aaagrtacgtrc tatrtrcaacc aaagcrtacc cagtaatrca aatrtaacatrc tacaacacrgc
360	agaaatrctgtr gcaatrccctc gcaagcrgtrc tccgtaatrca agaaatrctgaa tatrctgatrta
420	tgcraaatrcc ctaacgtratrca ttrgacatrta ataatccgtrgaa ataatrttrcga tctcaatcrtgtr
480	tcaaaagrgcrg agcatabatgtrc cactgtrctgtrta tgcrgaaatrca tgcgtaatrca tgcgtaatrca
540	ggcraacagggg ccaaaagggac agtatrtrgaac tatrtaaccrtcc tagggcrtccgaa aggggtrcacaac
600	aaacatrtrtrcc aaacccaa aaagaaagcrt tgcraacagrtat cgcctgaaatga ccaaacgaaag
660	taagtrctgtrc caagcatabtrtrc caaacgtrtrta ggcatabtrtrc aaagatrtrtaac acgtrgaaagaa
720	ttrgatrtrcraa ttrgtrtrtrcraa agtatrtaacrg atatrtaacrg cgaacgagtrcc ggcgtrgaaac
780	tgcrtgaaagaa gaatrtrtrtrcraa gtratrtrtrtrcraa cggcrttrtrtrcraa crrtrtrtrtrcraa aatrtrtrtrcraa
840	tccgaatrtrtrc caacagtrtrcraa ctcgacgaaagaa tcaatrtrtrtrcraa gatrtrtrtrtrcraa gatrtrtrtrcraa
900	gtrtrtrtrtrcraa ttrtrtrtrtrcraa tccgaaagtrta ctcraatrtaa tagtrtrtrtrcraa tatrtrtrtrcraa
960	ttrtrtrtrtrcraa ttrtrtrtrtrcraa acrtatrtrtrcraa cagtrtrtrtrcraa taaagtrtrtrcraa tcaatrtrtrtrcraa
1020	agtrtrtrtrtrcraa acrtagtrtrcraa aatrtrtrtrtrcraa ttrtrtrtrtrcraa ttrtrtrtrtrcraa gatrtrtrtrtrcraa
1080	tatrtrtrtrtrcraa aaagtrtrtrtrcraa cataaagggtrcraa tagatrtrtrtrcraa agtrtrtrtrtrcraa aagtrtrtrtrcraa
1140	aaagtrtrtrtrcraa gtrtrtrtrtrcraa aaagtrtrtrtrcraa cagtrtrtrtrtrcraa aatrtrtrtrtrcraa atcrtgtrtrtrcraa
1200	aaagtrtrtrtrcraa tatrtrtrtrtrcraa tatrtrtrtrtrcraa tatrtrtrtrtrcraa caaaaaagtrtrcraa gtrtrtrtrtrcraa
1260	tatrtrtrtrtrcraa acrtagtrtrtrcraa aaagtrtrtrtrcraa gtrtrtrtrtrcraa aaagtrtrtrtrcraa gtrtrtrtrtrcraa
1320	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1380	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1440	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1500	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1560	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1620	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1680	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1740	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1800	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1860	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1920	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
1980	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2040	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2100	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2160	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2220	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2280	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2340	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2400	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2460	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa
2520	aaacrtgtrtrtrcraa tatrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa aaacrtgtrtrtrcraa gtrtrtrtrtrcraa



-continued-

4980 aaagtgatgaa taatataatg tagatataac gatctctgacaa agtcttgagaa acttctcccg  
5040 tcatggttca gacctgtcaaa gagtgtatag gttctcaagag ttgatataag tctgtctccat  
5100 gaaatgatat catgtctctgaa agtcaatctc ttaaaagtg ttcaaatctat agaaagtgagg  
5160 tatgtttgct tagtctggtctc tgttgtgtcc ggtgagagtgga atttaccaga taatctgccc  
5220 ggtgtgtgtga gttgtctgcat gtttgacaaag agaatggaacaa gaggcggacga agccacacg  
5280 ggtcatatct aacctgtctg tcttaaaaag cgtttctagt ttaagtggt cccaatatac  
5340 ggtatataca caaaggtatgc agaaaagaaac atatggtcagg tctatgataa tattaataat  
5400 gtaaaaatga gttgctggtca cgtcccttctc tcatagaaat ttgtgtctctg gttatattgt  
5460 tataaaaata atataaata ggttttgagg gagaagtaaa cgtatgtgaa cgtatggaggaa  
5520 ccatgtgaaac tttctggaaag agttgtgtgat gagtctcatgag aatgtctcgtt  
5580 agactcgcga agtttctgaaac caatctctca aaaaaggtcc gaaaataata taatataat  
5640 ggttaaggggc gttcagggtcgg aagctctaaa ccaaaaagtt ttgatgaaagt tgaaaaagag  
5700 ttgtatataat tgatgaaagaa agtctcgtccg cgtatctgaa tctgataata  
5760 atatgtctca ctcaatcatc tctccatctc aatctgtgtc ttgtctcatc gttatgggtc  
5820 acctatagaa atgtttaaac gtttgttaaac atctgtatag taaaccttct caaacacagc  
5880 aagcaagaaac tactgttcaaa cagcagttca gctgaggtgtg gaaaaccttcc cctcagagca  
5940 cgtcagatc tccctgggtgat gtttataggg ttgtacaggtc caatctgcatc ttatgataccc  
6000 taatcatctgc gttctgtgggg gcttttgatga ctatgaaatag aatatactgaa gttgaaaacc  
6060 agcagagccc gacaaacagct gaacagttag atgtctacccc cagggtgatgac gatctcaggg  
6120 tctgaaatcc gttctgctata aatcaattag tcaatgaaat agtaagaggt acctgacctg  
6180 acaatcagaa tactttgaaa agtatgtctg ggttgtctg gacctctcga cctgcatctc  
6240 aatgtcagat gttctgtaaat ataaagtttg ttgttcttaaa acaacagttgg taccgtacgat  
6300 aagtaacagt gtttctccct ccaatcaat cgaaggttag ttctcttggag cgtcgtggagt  
6360 aaacatataat gttctcatata ttctcgttag caagtataaa agcgtgaggaa ttctgatatcc  
6384 cccggaaacc cgggtgggg ccca

<210 > SEQ ID NO 36  
<211 > LENGTH: 300  
<212 > TYPE: DNA  
<213 > ORGANISM: Escherichia coli

<400 > SEQUENCE: 36

60 caataacata cggctgtaca ttgcttcaaa atgctgtata gccccttga tgcctccatca  
120 cttctgatat atgtgacccca accttgcccgt aatgagtgtag cgcacatgaaa cgcagcagaa  
180 taagctgtcc ttgcccaccc ttctgtatata agacttccgg ctgatataccag acgttgcctcg  
240 catatataat atatactgca ttggtgaaat gatctgtataa actgctcctg acagcaatg  
300 cccgctcttc ttgtaaacgg ctttcccaac aacgctgatc aatcccaacag ttcttctgcat

<210 > SEQ ID NO 37  
<211 > LENGTH: 100  
<212 > TYPE: DNA  
<213 > ORGANISM: Escherichia coli

<400 > SEQUENCE: 37

60 tctgcaaac ttgatctgca actgctctgg acagcaatg cccgctcttc ttgtataacggc  
100 ctttcccaac aacgctgatc aatcccaacag ttcttctgcat

-continued

<210> SEQ ID NO 38  
 <211> LENGTH: 50  
 <212> TYPE: DNA  
 <213> ORGANISM: Escherichia coli

<400> SEQUENCE: 38

ttgtaacgcg ctttoccacc aacgctgata aattccacag ttttcgcat 50

<210> SEQ ID NO 39  
 <211> LENGTH: 20  
 <212> TYPE: DNA  
 <213> ORGANISM: Escherichia coli

<400> SEQUENCE: 39

aattccacag ttttcgcat 20

- The invention claimed is:
1. A polynucleotide comprising a tobamoviral nucleotide sequence, the tobamoviral nucleotide sequence comprising:
    - a first nucleotide sequence encoding a tobamoviral replication protein;
    - a second nucleotide sequence encoding a tobamoviral movement protein, the second nucleotide sequence being located downstream of the first nucleotide sequence; and
    - a linking site for linking with an exogenous nucleotide sequence encoding a polypeptide to be expressed, wherein the linking site is located downstream of the second nucleotide sequence, and wherein the second nucleotide sequence comprises a mutation relative to the wild-type nucleotide sequence encoding the tobamoviral movement protein, wherein the mutation comprises an insertion of 300 or more contiguous nucleotides, and wherein the insertion is a sequence other than an intron.
  2. The polynucleotide according to claim 1, wherein the second nucleotide sequence comprises the mutation at any position from 17 to 795 of the nucleotide sequence shown in SEQ ID NO: 20.
  3. The polynucleotide according to claim 1, wherein the first nucleotide sequence encoding a tobamoviral replication protein is:
    - (i) a polynucleotide encoding a polypeptide having at least 80% identity to SEQ ID NO: 1; or
    - (ii) a polynucleotide encoding a polypeptide having at least 80% identity to SEQ ID NO: 2.
  4. The polynucleotide according to claim 1, wherein the second nucleotide sequence is selected from the group consisting of polynucleotides having at least 80% identity to SEQ ID NO: 13 and a polynucleotide which hybridizes with a polynucleotide having a sequence that is complementary to SEQ ID NO: 13 under a stringent condition for hybridization.
  5. The polynucleotide according to claim 1, wherein the tobamoviral replication protein and the tobamoviral movement protein are derived from a tobacco mosaic virus or a tomato mosaic virus.
  6. A vector comprising the polynucleotide recited in claim 1.
  7. A plant comprising the polynucleotide recited in claim 1.
  8. A plant comprising the vector recited in claim 6.
  9. A transformant comprising the polynucleotide recited in claim 1.
  10. A transformant comprising the vector recited in claim 6.
  11. A method for producing a polypeptide, comprising: transforming or transfecting a plant with the polynucleotide recited in claim 1.
  12. A method for producing a polypeptide, comprising the step of: transforming a cell with the polynucleotide recited in claim 1.
  13. A kit for producing a polypeptide, comprising the polynucleotide recited in claim 1.
  14. A method for producing a polypeptide, comprising the step of: transforming or transfecting a plant with the vector recited in claim 6.
  15. A method for producing a polypeptide, comprising the step of: transforming a cell with the vector recited in claim 6.
  16. A kit for producing a polypeptide, comprising the vector recited in claim 6.
  17. A method for producing a polypeptide, comprising the step of: using the plant recited in claim 7.
  18. A method for producing a polypeptide, comprising the step of: using the transformant recited in claim 9.
  19. A kit for producing a polypeptide, comprising the plant recited in claim 7.
  20. A kit for producing a polypeptide, comprising the transformant recited in claim 9.
  21. A polynucleotide comprising a tobamoviral nucleotide sequence, the tobamoviral nucleotide sequence comprising:
    - a first nucleotide sequence encoding a tobamoviral replication protein;
    - a second nucleotide sequence encoding a tobamoviral movement protein, the second nucleotide sequence being located downstream of the first nucleotide sequence; and
    - a linking site for linking with an exogenous nucleotide sequence encoding a polypeptide to be expressed, wherein the linking site is located downstream of the second nucleotide sequence, and wherein the second nucleotide sequence comprises a mutation relative to the wild-type nucleotide sequence encoding the tobamoviral movement protein, wherein the mutation comprises an insertion of 1258 or more contiguous nucleotides.

\* \* \* \* \*