

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2020283367 B2**

(54) Title  
**PPR PROTEIN CAUSING LESS AGGREGATION AND USE OF THE SAME**

(51) International Patent Classification(s)  
**A01K 67/027** (2006.01) **C12N 1/21** (2006.01)  
**C07K 7/06** (2006.01) **C12N 5/10** (2006.01)  
**C07K 7/08** (2006.01) **C12N 15/10** (2006.01)  
**C07K 14/46** (2006.01) **C12N 15/12** (2006.01)  
**C07K 19/00** (2006.01) **C12N 15/63** (2006.01)  
**C12N 1/15** (2006.01) **C12Q 1/68** (2018.01)  
**C12N 1/19** (2006.01)

(21) Application No: **2020283367** (22) Date of Filing: **2020.05.29**

(87) WIPO No: **WO20/241877**

(30) Priority Data

|      |                    |      |                   |      |           |
|------|--------------------|------|-------------------|------|-----------|
| (31) | Number             | (32) | Date              | (33) | Country   |
|      | <b>2019-100553</b> |      | <b>2019.05.29</b> |      | <b>JP</b> |

(43) Publication Date: **2020.12.03**

(44) Accepted Journal Date: **2025.04.10**

(71) Applicant(s)  
**EditForce, Inc.;Kyushu University, National University Corporation**

(72) Inventor(s)  
**YAGI, Yusuke;IMAI, Takayoshi;TAMAI, Takayuki;NAKAMURA, Takahiro;TERAMOTO, Takamasa**

(74) Agent / Attorney  
**Pizzeys Patent and Trade Mark Attorneys Pty Ltd, PO Box 291, WODEN, ACT, 2606, AU**

(56) Related Art  
**TING BAN ET AL: JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 288, no. 44, 18 September 2013 (2013-09-18), US, pages 31540 - 31548, XP055456495, ISSN: 0021-9258, DOI: 10.1074/jbc.M113.496828**

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局

(43) 国際公開日  
2020年12月3日(03.12.2020)



(10) 国際公開番号

**WO 2020/241877 A1**

(51) 国際特許分類:

*A01K 67/027* (2006.01) *C12N 1/19* (2006.01)  
*C07K 7/06* (2006.01) *C12N 1/21* (2006.01)  
*C07K 7/08* (2006.01) *C12N 15/10* (2006.01)  
*C07K 14/46* (2006.01) *C12N 15/12* (2006.01)  
*C07K 19/00* (2006.01) *C12N 15/63* (2006.01)  
*C12N 5/10* (2006.01) *C12Q 1/68* (2018.01)  
*C12N 1/15* (2006.01)

(21) 国際出願番号: PCT/JP2020/021473

(22) 国際出願日: 2020年5月29日(29.05.2020)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:  
特願 2019-100553 2019年5月29日(29.05.2019) JP

(71) 出願人: エディットフォース株式会社  
(**EDITFORCE, INC.**) [JP/JP]; 〒8100001 福岡県  
福岡市中央区天神 1-9-17 Fukuoka  
(JP). 国立大学法人九州大学 (**KYUSHU  
UNIVERSITY, NATIONAL UNIVERSITY COR-  
PORATION**) [JP/JP]; 〒8190395 福岡県福岡市  
西区元岡 7 4 4 Fukuoka (JP).

(72) 発明者: 八木 祐介(**YAGI Yusuke**); 〒8100001 福岡県福岡市中央区天神一丁目 9 番 1 7 号 エ  
ディットフォース株式会社内 Fukuoka (JP).  
今井 崇喜(**IMAI Takayoshi**); 〒8100001 福岡県  
福岡市中央区天神一丁目 9 番 1 7 号 エデ  
ィットフォース株式会社内 Fukuoka (JP). 玉  
井 喬之(**TAMAI Takayuki**); 〒8100001 福岡県  
福岡市中央区天神一丁目 9 番 1 7 号 エデ  
ィットフォース株式会社内 Fukuoka (JP). 中村

崇裕(**NAKAMURA Takahiro**); 〒8190395 福岡  
県福岡市西区元岡 7 4 4 国立大学法人九州  
大学内 Fukuoka (JP). 寺本 岳大(**TERAMOTO  
Takamasa**); 〒8190395 福岡県福岡市西区元岡  
7 4 4 国立大学法人九州大学内 Fukuoka (JP).

(74) 代理人: 特許業務法人特許事務所サイクス(**SIKS  
& CO.**); 〒1040031 東京都中央区京橋一丁目 8  
番 7 号 京橋日殖ビル 8 階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保  
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ,  
BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,  
CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,  
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH,  
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,  
MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ,  
NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,  
QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,  
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,  
US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保  
護が可能): ARIPO (BW, GH, GM, KE, LR, LS,  
MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,  
ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ,  
TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ,  
DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT,  
LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS,  
SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM,  
GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** PPR PROTEIN WITH LESS AGGREGATION AND USE THEREOF

(54) 発明の名称: 凝集の少ない P P R タンパク質及びその利用

(57) **Abstract:** In order to improve the agglomerating property of PPR protein, A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N terminus is configured to be more hydrophilic. Further, A<sub>9</sub> amino acid of M<sub>1</sub> is configured to be a hydrophilic amino acid or glycine. A<sub>6</sub> amino acid is preferably asparagine or aspartic acid, and A<sub>9</sub> amino acid is preferably glutamine, glutamic acid, lysine, or glycine. Proteins including such PPR motifs as M<sub>1</sub> motif have not only an improved agglomerating property but also can have high affinity to a target nucleic acid.

(57) 要約: PPRタンパク質の凝集性を改善するために、N末端から1番目のPPRモチーフ(M<sub>1</sub>)のA<sub>6</sub>アミノ酸を、より親水性とする。さらにM<sub>1</sub>のA<sub>9</sub>アミノ酸を、親水性アミノ酸又はグリシンとする。A<sub>6</sub>アミノ酸は、好ましくはアスパラギン又はアスパラギン酸であり、A<sub>9</sub>アミノ酸は、好ましくはグルタミン、グルタミン酸、リジン、又はグリシンである。このようなPPRモチーフをM<sub>1</sub>モチーフとして含むタンパク質は、凝集性が改善されているのみならず、標的核酸に対する高い結合力を有しうる。



WO 2020/241877 A1

添付公開書類：

- 一 国際調査報告（条約第21条(3)）
- 一 明細書の別個の部分として表した配列リスト  
（規則5.2(a)）

## Description

Title of the Invention: PPR protein causing less aggregation and use of the same

### Technical Field

[0001]

The present invention relates to a nucleic acid manipulation technique using a protein capable of binding to an intended nucleic acid. The present invention is useful in a wide range of fields, including medicine (drug discovery support, therapeutic treatment etc.), agriculture (agricultural, fishery and livestock production, breeding etc.), and chemistry (biological material production etc.).

### Background Techniques

[0002]

PPR proteins are proteins comprising repeat of PPR motifs each having about 35 amino acids length, and one PPR motif can specifically bind to one base. The combination of the first, fourth, and ii-th (second from the end before the next motif) amino acids in a PPR motif determines to which one of adenine, cytosine, guanine, and uracil (or thymine) the motif binds (Patent documents 1 and 2).

[0003]

Among the naturally occurring RNA-binding PPR motifs, the combinations corresponding to each of the bases most frequently occurring are: first valine, fourth threonine, and ii-th asparagine for adenine; first valine, fourth asparagine, and ii-th serine for cytosine; first valine, fourth threonine, and ii-th aspartic acid for guanine, and first valine, fourth asparagine, and ii-th aspartic acid for uracil (Non-patent documents 1 to 5). By using these combinations of amino acids, PPR proteins that can specifically bind to an arbitrary sequence can be designed.

### Prior Art References

Patent documents

[0004]

Patent document 1: International Publication WO2013/058404

Patent document 2: International Publication WO2014/175284

Patent document 3: Japanese Patent Application No. 2019-100551

Non-patent documents

[0005]

Non-patent document 1: Coquille, S. et al., An artificial PPR scaffold for programmable RNA recognition, Nature Communications 5, Article number: 5729 (2014)

Non-patent document 2: Shen, C. et al., Specific RNA Recognition by Designer Pentatricopeptide Repeat Protein, Molecular Plant 8, 667-670 (2015)

Non-patent document 3: Shen, C. et al., Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins, Nature Communications, Volume 7, Article number: 11285 (2016)

Non-patent document 4: Miranda, R.G. et al., RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions, Nucleic Acids Research, 46(5), 2613-2623 (2018)

Non-patent document 5: Yan, J. et al., Delineation of pentatricopeptide repeat codes for target RNA prediction, Nucleic Acids Research, gkz075 (2019)

#### Summary of the Invention

Object to be achieved by the invention

[0006]

The inventors of the present invention have examined preparation of PPR proteins having high performance and formed by linking many, for example, 15 or more, of PPR motifs using the amino acid combinations mentioned above (Patent Document 3). On the other hand, according to the examination of the inventors of the present invention, it was found that some of the PPR proteins prepared in such a manner show aggregation property. In particular, when the PPR proteins were expressed in cultured animal cells, aggregation may be observed.

Means for achieving the object

[0007]

Therefore, the inventors of the present invention examined to solve this problem by amino acid mutation in the PPR motifs. Then, they found that the aggregation properties of PPR proteins can be improved by changing the 6th, preferably the 6th and 9th, amino acids of the first motif (on the N-terminal side) of the PPR protein to hydrophilic amino acids, and accomplished the present invention.

[0008]

The present invention provides the followings.

[1] A PPR motif, which is any one of the following PPR motifs:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a uracil-binding property; and

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a uracil-binding property.

[2] A PPR motif, which is any one of the following PPR motifs:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the

sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a uracil-binding property; and

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a uracil-binding property.

- a combination of asparagine as the amino acid at position 6 and glutamic acid as the amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and glutamine as the amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and lysine as the amino acid at position 9, and

- a combination of aspartic acid as the amino acid at position 6 and glycine as the

amino acid at position 9.

[3] The PPR motif according to 1 or 2, which is any one of the following PPR motifs:

(C-4) a PPR motif consisting of the sequence of SEQ ID NO: 4;

(A-4) a PPR motif consisting of the sequence of SEQ ID NO: 58;

(G-4) a PPR motif consisting of the sequence of SEQ ID NO: 59; and

(U-4) a PPR motif consisting of the sequence of SEQ ID NO: 60.

[4] Use of the PPR motif according to any one of 1 to 3 in a PPR protein as the first PPR motif from the N-terminus.

[5] The use according to 4, which is for reducing aggregation of the PPR protein.

[6] A protein capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises 1 to 30 of PPR motifs represented by the formula 1 mentioned below, and wherein the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus is a hydrophilic amino acid:

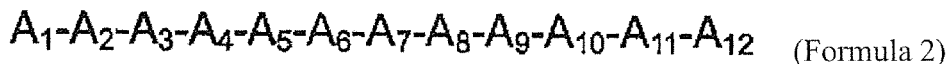
[Formula 1]



wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

[Formula 2]



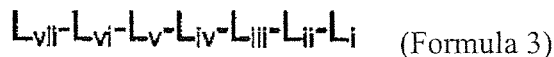
wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

[Formula 3]



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist.

[7] The protein according to 6, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is a hydrophilic amino acid or glycine.

[8] The protein according to 6 or 7, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> is asparagine or

aspartic acid.

[9] The protein according to any one of 6 to 8, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is glutamine, glutamic acid, lysine, or glycine.

[10] The protein according to any one of 6 to 9, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> and the A<sub>9</sub> amino acid of M<sub>1</sub> correspond to any of the following combinations:

- combination of asparagine as the A<sub>6</sub> amino acid and glutamic acid as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and glutamine as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and lysine as the A<sub>9</sub> amino acid, and
- combination of aspartic acid as the A<sub>6</sub> amino acid and glycine as the A<sub>9</sub> amino acid.

[11] A fusion protein of at least one selected from the group consisting of a fluorescent protein, a nuclear localization signal peptide, and a tag protein, and a PPR protein containing the PPR motif according to any one of 1 to 3 as the first PPR motif from the N-terminus, or the protein according to any one of 6 to 10.

[12] A method for modifying a PPR protein containing the PPR motif according to 6 and capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises making the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus more hydrophilic.

[13] A method for detecting a nucleic acid, which uses a PPR protein containing the PPR motif according to any one of 1 to 3 as the first PPR motif from the N-terminus, the protein according to any one of 6 to 10, or the fusion protein according to 11.

[14] A nucleic acid encoding the PPR motif according to any one of 1 to 3, a PPR protein containing the PPR motif according to any one of 1 to 3 as the first PPR motif from the N-terminus, or the protein according to any one of 6 to 10.

[15] A vector comprising the nucleic acid according to 14.

[16] A cell (except for human individual) containing the vector according to 15.

[17] A method for manipulating a nucleic acid, which uses the PPR motif according to any one of 1 to 3, a PPR protein containing the PPR motif according to any one of 1 to 3 as the first PPR motif from the N-terminus, the protein according to any one of 6 to 10, or the vector according to 15 (implementation in human individual is excluded).

[18] A method for producing an organism, which comprises the manipulation method according to 17.

[0009]

The present invention also provides the followings.

[1] A PPR motif, which is any one of the following PPR motifs:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a uracil-binding property; and

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a uracil-binding property.

[2] A PPR motif, which is any one of the following PPR motifs:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the

sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a uracil-binding property; and

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a uracil-binding property.

- a combination of asparagine as the amino acid at position 6 and glutamic acid as the amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and glutamine as the amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and lysine as the amino acid at position 9, and

- a combination of aspartic acid as the amino acid at position 6 and glycine as the

amino acid at position 9.

[3] Use of the PPR motif according to 1 or 2 in a PPR protein as the first PPR motif from the N-terminus.

[4] The use according to 3, which is for reducing aggregation of the PPR protein.

[5] A protein capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises 1 to 30 of PPR motifs represented by the formula 1 mentioned below, and wherein the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus is a hydrophilic amino acid:

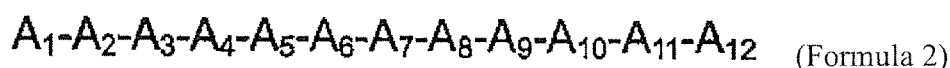
[Formula 4]



wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

[Formula 5]



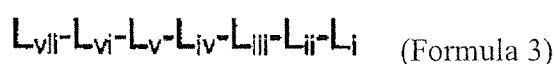
wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

[Formula 6]



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist.

[6] The protein according to 5, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is a hydrophilic amino acid or glycine.

[7] The protein according to 5 or 6, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> is asparagine or aspartic acid.

[8] The protein according to any one of 5 to 7, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is glutamine, glutamic acid, lysine, or glycine.

[9] The protein according to any one of 5 to 8, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> and the A<sub>9</sub> amino acid of M<sub>1</sub> correspond to any of the following combinations:

- combination of asparagine as the A<sub>6</sub> amino acid and glutamic acid as the A<sub>9</sub> amino acid,
  - combination of asparagine as the A<sub>6</sub> amino acid and glutamine as the A<sub>9</sub> amino acid,
  - combination of asparagine as the A<sub>6</sub> amino acid and lysine as the A<sub>9</sub> amino acid, and
  - combination of aspartic acid as the A<sub>6</sub> amino acid and glycine as the A<sub>9</sub> amino acid.
- [10] A fusion protein of at least one selected from the group consisting of a fluorescent protein, a nuclear localization signal peptide, and a tag protein, and a PPR protein containing the PPR motif according to 1 or 2 as the first PPR motif from the N-terminus, or the protein according to any one of 5 to 9.
- [11] A method for modifying a PPR protein containing the PPR motif according to 3 and capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises making the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus more hydrophilic.
- [12] A method for detecting a nucleic acid, which uses a PPR protein containing the PPR motif according to 1 or 2 as the first PPR motif from the N-terminus, the protein according to any one of 5 to 9, or the fusion protein according to 10.
- [13] A nucleic acid encoding the PPR motif according to 1 or 2, a PPR protein containing the PPR motif according to 1 or 2 as the first PPR motif from the N-terminus, or the protein according to any one of 5 to 9.
- [14] A vector comprising the nucleic acid according to 13.
- [15] A cell (except for human individual) containing the vector according to 14.
- [16] A method for manipulating a nucleic acid, which uses the PPR motif according to 1 or 2, a PPR protein containing the PPR motif according to 1 or 2 as the first PPR motif from the N-terminus, the protein according to any one of 5 to 9, or the vector according to 14 (implementation in human individual is excluded).
- [17] A method for producing an organism, which comprises the manipulation method according to 16.

#### Brief Description of the Drawings

[0010]

[Fig. 1] The method for designing a PPR motif. A: The 6th and 9th amino acids of the first motif are exposed to the outside. B: For the 6th and 9th amino acids of the first motif that recognize cytosine, there were chosen leucine and glycine (C<sub>6</sub>L9G) as a typical combination, and leucine and glutamic acid (C<sub>6</sub>L9E), asparagine and glutamine (C<sub>6</sub>N9Q), asparagine and glutamic acid (C<sub>6</sub>N9E), asparagine and lysine (C<sub>6</sub>N9K), and aspartic acid and glycine (C<sub>6</sub>D9G) as mutant types.

[Fig. 2] Aggregation and localization into nuclei of each PPR protein. Intracellular expression of PPR fused with GFP and a nuclear localization signal sequence was confirmed on the basis of fluorescence microscopy images. When the protein was fused with EGFP, it was observed that PPRcag\_1 (6L9G) and PPRcag\_2 (6L9E) did not localize to the nuclei, but strongly aggregated around the nuclei. On the other hand, PPRcag\_3 (6N9Q), PPRcag\_4 (6N9E), PPRcag\_5 (6N9K), and PPRcag\_6 (6D9G) did not localize to the nuclei, although their aggregation was weak. When mClover3 was fused, it was observed that PPRcag\_1 (6L9G) and PPRcag\_2 (6L9E) localized to the nuclei, but they aggregated in the nuclei. PPRcag\_3 (6N9Q), PPRcag\_4 (6N9E), PPRcag\_5 (6N9K), and PPRcag\_6 (6D9G) localized to the nuclei, and showed no aggregation.

[Fig. 3] Binding experiment of PPR protein and RNA. It was found that all PPR proteins including amino acid mutations of the 6th and 9th amino acids specifically bind to the target CAGx6. In comparison with PPRcag\_1, the binding power to the target sequence was substantially the same for PPRcag\_2, about 80% for PPRcag\_3, about 60% for PPRcag\_4, about 120% for PPRcag\_5, and about 130% for PPRcag\_6.

[Fig. 4] The effect of the first PPR motif from the N-terminus on aggregation. Each PPR protein was prepared in an *E. coli* expression system, purified, and separated by gel filtration chromatography. The smaller volume of the elution fraction (Elution vol.) means a larger molecular size. The proteins using v2 were eluted in 8 to 10 mL of elution fractions, whereas the peaks of the proteins using v3.2 were observed in elution fractions of 12 to 14 mL. These results suggested possibility that the proteins using v2 aggregated due to the larger protein size thereof, and it was found that the aggregation was improved in the proteins using v3.2.

## Modes for Carrying out the Invention

[0011]

[PPR motif and PPR protein]

(Definition)

The PPR motif referred to in the present invention means a polypeptide constituted by 30 to 38 amino acids and having an amino acid sequence of an E value not larger than a predetermined value (desirably E-03) obtained for PF01535 in Pfam or PS51375 in Prosite as determined by amino acid sequence analysis with a protein domain search program on the Web, unless especially stated. The position numbers of amino acids constituting the PPR motif defined in the present invention are substantially synonymous with those of PF01535, and they correspond to those

obtained by subtracting 2 from the numbers of the amino acid positions of PS51375 (for example, the position 1 referred to in the present invention corresponds to the position 3 of PS51375). Further, the term "ii" (-2)-th amino acid means the second amino acid from the end (C-terminus side) of the amino acids constituting the PPR motif, or the second amino acid towards the N-terminus side from the first amino acid of the following PPR motif, i.e., -2nd amino acid. When the following PPR motif is not definitely identified, the amino acid 2 amino acids before the first amino acid of the following helical structure is the amino acid of "ii". For Pfam, <http://pfam.sanger.ac.uk/> can be referred to, and for Prosite, <http://www.expasy.org/prosite/> can be referred to.

[0012]

Although the conservativeness of the conserved amino acid sequence of the PPR motif is low at the amino acid level, two of the  $\alpha$ -helices as the secondary structure are well conserved. Although a typical PPR motif is constituted by 35 amino acids, the length thereof is as variable as is from 30 to 38 amino acids.

[0013]

More specifically, the PPR motif referred to in the present invention consists of a polypeptide of a 30- to 38-amino acid length represented by the formula 1.

[0014]

[Formula 7]

**(Helix A)-X-(Helix B)-L** (Formula 1)

wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

[0015]

[Formula 8]

**A<sub>1</sub>-A<sub>2</sub>-A<sub>3</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>-A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>** (Formula 2)

wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

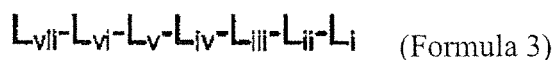
X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

[0016]

[Formula 9]



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that  $L_{iii}$  to  $L_{vii}$  may not exist.

[0017]

The term PPR protein used in the present invention refers to a PPR protein comprising one or more, preferably two or more, of the above-mentioned PPR motifs, unless especially indicated. The term protein used in this description refers to any substance consisting of a polypeptide (chain consisting of a plurality of amino acids bound via peptide bonds), unless especially indicated, and includes those consisting of a polypeptide of a comparatively low molecular weight. The term amino acid used in the present invention refers to a usual amino acid molecule, and also refers to an amino acid residue constituting a peptide chain. Which one is referred to shall be clear to those skilled in the art from the context.

[0018]

In the present invention, the term specificity/specific used for the binding property of the PPR motif to a base in the target nucleic acid means that the binding activity to any one of the four bases is higher than the binding activities to the other bases, unless especially stated.

[0019]

In the present invention, the term nucleic acid refers to RNA or DNA. Although the PPR protein may have specificity for bases in RNA or DNA, it does not bind to nucleic acid monomers.

[0020]

In the PPR motif, combination of three of the 1st, 4th, and ii-th amino acids is important for specific binding to a base, and to which base the motif binds can be determined according to this combination (Patent document 1 and 2 mentioned above).

[0021]

Specifically, with respect to the RNA-binding PPR motifs, the relationship between the combinations of three of the 1st, 4th, and ii-th amino acids and the bases to which they can bind is as follows (see Patent document 1 mentioned above).

(3-1) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to U, less strongly to C, and still less strongly to A or G.

(3-2) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to A, less strongly to G, and still less strongly to C, but does not bind to U.

(3-3) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to C, and less strongly to A or U, but does not bind to G.

(3-4) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glutamic acid, glycine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to G, but does not bind to A, U, and C.

(3-5) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to C, less strongly to U, and still less strongly to A, but does not bind to G.

(3-6) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to G, and less strongly to U, but does not bind to A and C.

(3-7) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of lysine, threonine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to G, and less strongly to A, but does not bind to U and C.

(3-8) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, serine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to A, less strongly to C, and still less strongly to G and U.

(3-9) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and serine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to C, and less strongly to U, but does not bind to A and G.

(3-10) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, threonine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to A, but does not bind to G, U, and C.

(3-11) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to U, and less strongly to A, but does not bind to G and C.

(3-12) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of threonine, threonine, and asparagine in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to A, but does not bind to G, U, and C.

(3-13) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, methionine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to U, and less strongly to C, but does not bind to A and G.

(3-14) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, proline, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to U, and less strongly to C, but does not bind to A and G.

(3-15) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, proline, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to U, but does not bind to A, G, and C.

(3-16) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of leucine, threonine, and aspartic acid in this order, the PPR motif has such a selective RNA base-binding ability that the motif strongly binds to G, but does not bind to A, U, and C.

[0022]

Specifically, with respect to the DNA-binding PPR motifs, the relationship between combinations of the three of the 1st, 4th, and ii-th amino acids and the bases to which they can bind is as follows (see Patent document 2 mentioned above).

(2-1) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, glycine, and aspartic acid in this order, the PPR motif selectively binds to G.

(2-2) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glutamic acid, glycine, and aspartic acid in this order, the PPR motif selectively binds to G.

(2-3) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, glycine, and asparagine in this order, the PPR motif selectively

binds to A.

(2-4) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glutamic acid, glycine, and asparagine in this order, the PPR motif selectively binds to A.

(2-5) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, glycine, and serine in this order, the PPR motif selectively binds to A, and less selectively to C.

(2-6) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, isoleucine, and an arbitrary amino acid in this order, the PPR motif selectively binds to T and C.

(2-7) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, isoleucine, and asparagine in this order, the PPR motif selectively binds to T, and less selectively to C.

(2-8) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and an arbitrary amino acid in this order, the PPR motif selectively binds to T and C.

(2-9) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and aspartic acid in this order, the PPR motif selectively binds to C.

(2-10) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and lysine in this order, the PPR motif selectively binds to T.

(2-11) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, methionine, and an arbitrary amino acid in this order, the PPR motif selectively binds to T.

(2-12) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, methionine, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-13) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, methionine, and aspartic acid in this order, the PPR motif selectively binds to T, and less selectively to C.

(2-14) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and an arbitrary amino acid in this order, the PPR motif selectively binds to C and T.

(2-15) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and aspartic acid in this order, the PPR motif

selectively binds to T.

(2-16) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-17) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glycine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-18) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-19) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of threonine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-20) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T, and less selectively to C.

(2-21) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, asparagine, and aspartic acid in this order, the PPR motif selectively binds to T, and less selectively to C.

(2-22) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and asparagine in this order, the PPR motif selectively binds to C.

(2-23) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and asparagine in this order, the PPR motif selectively binds to C.

(2-24) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of serine, asparagine, and asparagine in this order, the PPR motif selectively binds to C.

(2-25) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and asparagine in this order, the PPR motif selectively binds to C.

(2-26) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and serine in this order, the PPR motif selectively binds to C.

(2-27) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and serine in this order, the PPR motif selectively binds to C.

(2-28) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and threonine in this order, the PPR motif selectively

binds to C.

(2-29) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and threonine in this order, the PPR motif selectively binds to C.

(2-30) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and tryptophan in this order, the PPR motif selectively binds to C, and less selectively to T.

(2-31) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and tryptophan in this order, the PPR motif selectively binds to T, and less selectively to C.

(2-32) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and an arbitrary amino acid in this order, the PPR motif selectively binds to T.

(2-33) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-34) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, proline, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-35) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, proline, and aspartic acid in this order, the PPR motif selectively binds to T.

(2-36) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and an arbitrary amino acid in this order, the PPR motif selectively binds to A and G.

(2-37) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and asparagine in this order, the PPR motif selectively binds to A.

(2-38) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, serine, and asparagine in this order, the PPR motif selectively binds to A.

(2-39) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, serine, and asparagine in this order, the PPR motif selectively binds to A.

(2-40) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and an arbitrary amino acid in this order, the PPR motif selectively binds to A and G.

(2-41) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and aspartic acid in this order, the PPR motif

selectively binds to G.

(2-42) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and aspartic acid in this order, the PPR motif selectively binds to G.

(2-43) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and asparagine in this order, the PPR motif selectively binds to A.

(2-44) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, threonine, and asparagine in this order, the PPR motif selectively binds to A.

(2-45) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, threonine, and asparagine in this order, the PPR motif selectively binds to A.

(2-46) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and asparagine in this order, the PPR motif selectively binds to A.

(2-47) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, valine, and an arbitrary amino acid in this order, the PPR motif binds to A, C, and T, but does not bind to G.

(2-48) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, valine, and aspartic acid in this order, the PPR motif selectively binds to C, and less selectively to A.

(2-49) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, valine, and glycine in this order, the PPR motif selectively binds to C.

(2-50) When the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, valine, and threonine in this order, the PPR motif selectively binds to T.

[0023]

(Particularly preferred combinations of the three amino acids)

For the RNA-binding PPR motifs, there are typical combinations of the 1st, 4th, and ii-th amino acids that can recognize and specifically bind to each base. Specifically, the combination that recognizes adenine consists of 1st valine, 4th threonine, and ii-th asparagine; the combination that recognizes cytosine consists of 1st valine, 4th asparagine, and ii-th serine; the combination that recognizes guanine consists of 1st valine, 4th threonine, and ii-th aspartic acid, and the combination that recognizes uracil consists of 1st is valine, 4th is asparagine, and ii-th aspartic acid (Non-patent documents 1 to 5 mentioned above). In one of the preferred

embodiments of the present invention, these combinations are used.

[0024]

(Improvement of aggregation property)

The inventors of the present invention found that the amino acid at position 6 of the PPR motif is extremely frequently hydrophobic amino acid (especially leucine) and the amino acid at position 9 is extremely frequently a non-hydrophilic amino acid (especially glycine) on the basis of the amino acid information of existing naturally occurring PPR motifs. On the basis of structures of the PPR proteins for which crystal structures have already been obtained (Non-patent document 6: Coquille et al., 2014 Nat. Commun., PDB ID: 4PJQ, 4WN4, 4WSL, 4PJR, Non-patent document 7: Shen et al., 2015 Nat. Commun., PDB ID: 5I9D, 5I9F, 5I9G, 5I9H), they imagined that since those 6th and 9th amino acids in the first motif (N-terminus side) are exposed to the outside, the proteins show aggregation property due to these exposed hydrophobic amino acids (Fig. 1A). On the other hand, they considered that, in the second and following motifs, the 6th and 9th amino acids are buried inside the protein, and form a hydrophobic core, and therefore if hydrophilic residues are placed as the 6th and 9th amino acids of all the motifs, the protein structure may collapse. Therefore, they decided to decrease the aggregation property of PPR by using hydrophilic amino acid (asparagine, aspartic acid, glutamine, glutamic acid, lysine, arginine, serine, and threonine) as the 6th amino acid, preferably the 6th and 9th amino acids, in only the first motif.

[0025]

Specific procedure is as follows.

In the first PPR motif ( $M_1$ ) from the N-terminus of a protein capable of binding to a target nucleic acid having a specific nucleotide sequence:

- (1) a hydrophilic amino acid is used as the  $A_6$  amino acid, preferably asparagine or aspartic acid is used as the  $A_6$  amino acid,
- (2) further, a hydrophilic amino acid or glycine, preferably glutamine, glutamic acid, lysine, or glycine, is used as the  $A_9$  amino acid, or
- (3) the  $A_6$  amino acid and  $A_9$  amino acid are constituted by any of the following combinations;
  - combination of asparagine as the  $A_6$  amino acid and glutamic acid as the  $A_9$  amino acid,
  - combination of asparagine as the  $A_6$  amino acid and glutamine as the  $A_9$  amino acid,
  - combination of asparagine as the  $A_6$  amino acid and lysine as the  $A_9$  amino acid, and
  - combination of aspartic acid as the  $A_6$  amino acid and glycine as the  $A_9$  amino acid.

[0026]

(Novel PPR motif)

The present invention provides novel PPR motifs with improved aggregation property and novel PPR proteins containing the same, which were found as described above.

[0027]

The novel PPR motifs provided by the present invention are followings:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, and having a uracil-binding property; and

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1),

provided that the amino acids at positions 1, 4, 6, and 34 are identical, and having a uracil-binding property.

[0028]

Among such PPR motifs, the followings are particularly preferred:

(C-1) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7;

(C-2) a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a cytosine-binding property;

(C-3) a PPR motif having a sequence identity of at least 80% to any one of the sequences of SEQ ID NOS: 4 to 7, provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a cytosine-binding property;

(A-1) a PPR motif consisting of the sequence of SEQ ID NO: 8 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(A-2) a PPR motif consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having an adenine-binding property;

(A-3) a PPR motif having a sequence identity of at least 80% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having an adenine-binding property;

(G-1) a PPR motif consisting of the sequence of SEQ ID NO: 9 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(G-2) a PPR motif consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a guanine-binding property;

(G-3) a PPR motif having a sequence identity of at least 80% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a guanine-binding property;

(U-1) a PPR motif consisting of the sequence of SEQ ID NO: 10 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(U-2) a PPR motif consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, and having a uracil-binding property;

(U-3) a PPR motif having a sequence identity of at least 80% to the sequence of (U-1),

provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, and having a uracil-binding property:

- a combination of asparagine as the amino acid at position 6 and glutamic acid as the amino acid at position 9,
- a combination of asparagine as the amino acid at position 6 and glutamine as the amino acid at position 9,
- a combination of asparagine as the amino acid at position 6 and lysine as the amino acid at position 9, and
- a combination of aspartic acid as the amino acid at position 6 and glycine as the amino acid at position 9.

[0029]

The specific sequences of SEQ ID NOS: 4 to 10 are shown in Fig. 1, and in the sequence listing.

[0030]

Among such PPR motifs, more preferred are the followings:

- (C-4) a PPR motif consisting of the sequence of SEQ ID NO: 4;
- (A-4) a PPR motif consisting of the sequence of SEQ ID NO: 58;
- (G-4) a PPR motif consisting of the sequence of SEQ ID NO: 59; and
- (U-4) a PPR motif consisting of the sequence of SEQ ID NO: 60.

[0031]

The sequences of SEQ ID NOS: 58 to 60 are shown below and in the sequence listing.

Sequence of SEQ ID NO: 58:

VTYTTNIDQLCKAGKVDEALELFKEMRSKGVKPNV

Sequence of SEQ ID NO: 59:

VTYTTNIDQLCKAGKVDEALELFDKMKERGIKPDV

Sequence of SEQ ID NO: 60:

VTYNTNIDQLCKAGRLDEAEELLEEMEEKGIKPDV

[0032]

(PPR protein with improved aggregation property)

The present invention also provides PPR proteins with improved aggregation properties found as described above.

[0033]

In one of the preferred embodiments, the A<sub>9</sub> amino acid of M<sub>1</sub> is a non-hydrophobic amino acid or glycine, whatever the other amino acids of M<sub>1</sub> are, and whatever the amino acid sequences of the motifs other than M<sub>1</sub> are. The

non-hydrophobic amino acid is a hydrophilic amino acid, or cysteine or histidine; preferably a hydrophilic amino acid, i.e., arginine, asparagine, aspartic acid, glutamic acid, glutamine, lysine, serine, or threonine; more preferably glutamine, glutamic acid, or lysine.

[0034]

In one of the preferred embodiments, the A<sub>9</sub> amino acid of M<sub>1</sub> is glutamine, glutamic acid, lysine, or glycine, whatever the other amino acids of M<sub>1</sub> are, and whatever the amino acid sequences of the motifs other than M<sub>1</sub> are.

[0035]

In one of the preferred embodiments, the A<sub>6</sub> amino acid of M<sub>1</sub> is a non-hydrophobic amino acid, whatever the other amino acids of M<sub>1</sub> are, and whatever the amino acid sequences of the motifs other than M<sub>1</sub> are. The non-hydrophobic amino acid is, for example, a hydrophilic amino acid, or cysteine or histidine; preferably a hydrophilic amino acid, i.e., arginine, asparagine, aspartic acid, glutamic acid, glutamine, lysine, serine, or threonine; more preferably asparagine, or aspartic acid.

[0036]

In one of the particularly preferred embodiments, the A<sub>6</sub> and A<sub>9</sub> amino acids of M<sub>1</sub> consist of any of the following combination, whatever the other amino acids of M<sub>1</sub> are, and whatever the amino acid sequences of the motifs other than M<sub>1</sub> are:

- combination of asparagine as the A<sub>6</sub> amino acid and glutamic acid as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and glutamine as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and lysine as the A<sub>9</sub> amino acid, and
- combination of aspartic acid as the A<sub>6</sub> amino acid and glycine as the A<sub>9</sub> amino acid.

[0037]

In one of the preferred embodiments of the RNA-binding protein, the A<sub>6</sub> and A<sub>9</sub> amino acids of M<sub>1</sub> satisfy the above conditions, and at least one, preferably half or more, more preferably all, of the included PPR motifs satisfy any of the following conditions:

- when the base to be bound is cytosine, A<sub>1</sub> is valine, A<sub>4</sub> is asparagine, and A<sub>ii</sub> is serine;
- when the base to be bound is adenine, A<sub>1</sub> is valine, A<sub>4</sub> is threonine, and A<sub>ii</sub> is asparagine;
- when the base to be bonded is guanine, A<sub>1</sub> is valine, A<sub>4</sub> is threonine, and A<sub>ii</sub> is aspartic acid; and
- when the base to be bound is uracil or thymine, A<sub>1</sub> is valine, A<sub>4</sub> is asparagine, and A<sub>ii</sub>

is aspartic acid.

[0038]

In one of the preferred embodiments of the RNA-binding protein, M<sub>1</sub> is the novel PPR motif described above.

[0039]

In one of the particularly preferred embodiments, M<sub>1</sub> is a PPR motif consisting of any of the following polypeptides:

- a polypeptide consisting of any one of the sequences of SEQ ID NOS: 4 to 7 for cytosine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 8 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined in the following paragraph is satisfied for adenine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 9 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined in the following paragraph is satisfied for guanine as the base to be bound; and
- a polypeptide consisting of the sequence of SEQ ID NO: 10 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined in the following paragraph is satisfied for uracil as the base to be bound;

At least one of the PPR motifs other than M<sub>1</sub> is a PPR motif consisting of any one of the following polypeptides:

- a polypeptide consisting of the sequence of SEQ ID NO: 2 for cytosine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 8 for adenine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 9 for guanine as the base to be bound; and
- a polypeptide consisting of the sequence of SEQ ID NO: 10 for uracil as the base to be bound.

[0040]

The combinations referred to in the above paragraph are any of the followings:

- a combination of asparagine as the A<sub>6</sub> amino acid and glutamic acid as the A<sub>9</sub> amino acid at position 9,
- a combination of asparagine as the A<sub>6</sub> amino acid and glutamine as the A<sub>9</sub> amino acid,
- a combination of asparagine as the A<sub>6</sub> amino acid and lysine as the A<sub>9</sub> amino acid,

and

- a combination of aspartic acid as the A<sub>6</sub> amino acid and glycine as the A<sub>9</sub> amino acid.  
[0041]

In one of the particularly preferred embodiments, M<sub>1</sub> is a PPR motif consisting of any one of the following polypeptides:

- a polypeptide consisting of the sequence of SEQ ID NO: 4 for cytosine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 58 for adenine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 59 for guanine as the base to be bound; and
- a polypeptide consisting of the sequence of SEQ ID NO: 60 for uracil as the base to be bound.

At least one of the PPR motifs other than M<sub>1</sub> is a PPR motif consisting of any one of the following polypeptides:

- a polypeptide consisting of the sequence of SEQ ID NO: 2 for cytosine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 8 having a substitution of the amino acid at position 15 with lysine for adenine as the base to be bound;
- a polypeptide consisting of the sequence of SEQ ID NO: 9 for guanine as the base to be bound; and
- a polypeptide consisting of the sequence of SEQ ID NO: 10 for uracil as the base to be bound.

[0042]

(Use of skeleton of high performance PPR motif)

In one of the preferred embodiments of the present invention, the amino acids in the PPR motifs for cytosine, adenine, guanine, and uracil (or thymine) other than the amino acids at positions 1, 4, 6, 9 and ii can be particular amino acids. Precisely, from the *Arabidopsis thaliana* PPR motif sequences, there were collected the PPR motifs in which the combination of amino acids at positions 1, 4, and ii is VTN as adenine-recognizing PPR motifs, those in which the same is VSN as the cytosine-recognizing PPR motifs, those in which the same is VTD as the guanine-recognizing PPR motifs, and those in which the same is VND as the uracil-recognizing PPR motifs, and types and occurring numbers of the amino acids at every position are summarized. Then, by selecting amino acid highly frequently occurring at each position, the performance of the PPR motif can be enhanced.

[0043]

For the purpose of using highly frequently occurring amino acids as the amino acids other than the 1st, 4th, 6th, 9th and ii-th amino acids as described above, for obtaining an RNA-binding PPR protein, the amino acid sequences of the following PPR motifs can be referred to:

a PPR motif consisting of any one of the sequences of SEQ ID NOS: 4 to 7 as a PPR motif for cytosine;

a PPR motif consisting of the sequence of SEQ ID NO: 8 as the PPR motif for adenine;

a PPR motif consisting of the sequence of SEQ ID NO: 9 as the PPR motif for guanine; and

a PPR motif consisting of the sequence of SEQ ID NO: 10 as the PPR motif for guanine;

[0044]

(Explanation of technical terms, etc.)

The term "identity" used in the present invention for base sequence (also referred to as nucleotide sequence) or amino acid sequence means percentage of number of matched bases or amino acids shared between two sequences aligned in an optimal manner, unless especially stated. In other words, the identity can be calculated in accordance with the equation:  $\text{Identity} = (\text{Number of matched positions} / \text{Total number of positions}) \times 100$ , and it can be calculated by using commercially available algorithms. Such algorithms are also incorporated in the NBLAST and XBLAST programs described in Altschul et al., J. Mol. Biol., 215 (1990) 403-410. In more detail, the search and analysis for the identity of nucleotide or amino acid sequences can be performed with algorithms or programs well known to those skilled in the art (e.g., BLASTN, BLASTP, BLASTX, and ClustalW). In the case of using a program, parameters can be appropriately set by those skilled in the art, or the default parameters of each program can also be used. The specific procedures of these analysis methods are also well known to those skilled in the art.

[0045]

In this description, when the expression of having an identity (or identity is high) is used for a nucleotide sequence or amino acid sequence, it means for both cases to have an identity of, at least 70%, preferably 80% or higher, more preferably 85% or higher, still more preferably 90% or higher, further preferably 95% or higher, still further preferably 97.5% or higher, even more preferably 99% or higher, unless especially stated.

[0046]

As for the term "sequence having a substitution, deletion, or addition" used in the present invention concerning PPR motif or protein, the number of amino acids substituted or the like is not particularly limited in any motif or protein, so long as the motif or protein consisting of the amino acid sequence has the desired function, unless especially stated. The number of amino acids to be substituted, or the like may be about 1 to 9 or 1 to 4, or even larger number of amino acids may be substituted or the like if they are substituted with amino acids having similar properties. The means for preparing polynucleotides or proteins for such amino acid sequences are well known to those skilled in the art.

[0047]

Amino acids having similar properties refer to amino acids with similar physical properties such as hydropathy, charge, pKa, and solubility, and refer to such amino acid as mentioned below, for example.

Hydrophobic amino acids; alanine, valine, glycine, isoleucine, leucine, phenylalanine, proline, tryptophan, tyrosine.

Non-hydrophobic amino acids; arginine, asparagine, aspartic acid, glutamic acid, glutamine, lysine, serine, threonine, cysteine, histidine, methionine.

Hydrophilic amino acids; arginine, asparagine, aspartic acid, glutamic acid, glutamine, lysine, serine, threonine.

Acidic amino acids: aspartic acid, glutamic acid.

Basic amino acids: lysine, arginine, histidine.

Neutral amino acids: alanine, asparagine, cysteine, glutamine, glycine, isoleucine, leucine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine.

Sulfur-containing amino acids: methionine, cysteine.

Aromatic ring-containing amino acids: tyrosine, tryptophan, phenylalanine.

[0048]

The term "preparation" used for genes, nucleic acids, polynucleotides, proteins, motifs, etc. can be rephrased as "production" or "manufacturing". In addition, the term "construction" is sometimes used to refer to preparation of genes or the like by combining parts, and "construction" can also be rephrased as "production" or "manufacturing".

[0049]

The PPR motif, protein containing the same, or nucleic acids encoding the same of the present invention can be prepared by those skilled in the art using

conventional techniques, and the explanations in the section of Examples of this description.

[0050]

[Characteristics and use of PPR proteins]

(Improvement of aggregation property of PPR proteins)

The PPR proteins produced by using the novel PPR motifs of the present invention show reduced aggregation in cells. Aggregation of a PPR protein can be evaluated by those skilled in the art by expressing the PPR protein in cells and confirming presence or absence of aggregation. The confirmation is more easily performed by fusing the PPR protein to a fluorescent protein and expressing them. According to the examination of the inventors of the present invention, by appropriately modifying the amino acids in the 1st motif of the PPR protein, the aggregation property of the PPR protein in the cells is improved, and localization of the same to the nuclei is improved.

[0051]

(Binding power)

PPR proteins prepared by using the novel PPR motifs of the present invention may have not only reduced intracellular aggregation property, but also RNA binding performance equivalent to or higher than those of PPR proteins for the same target RNA prepared by using existing PPR motifs. Equivalent means to be 55% or higher, preferably about 75%.

[0052]

The binding power to a target sequence can be evaluated by EMSA (Electrophoretic Mobility Shift Assay) or a method using Biacore. EMSA is a method utilizing a property of nucleic acid that when a sample consisting of a nucleic acid bound with a protein is electrophoresed, the mobility of the nucleic acid molecule changes from that of the nucleic acid not bound. Molecular interaction analyzers, such as Biacore as a typical example, enable kinetic analysis, and therefore allow detailed protein-nucleic acid binding analysis.

[0053]

The binding power to a target sequence can also be evaluated by adding a solution containing a candidate protein to a solid-phased target nucleic acid, and detecting or quantifying the protein that bound to the target nucleic acid. This method is sometimes referred to as the RPB-ELISA (RNA-protein binding ELISA) method, since it utilizes ELISA (Enzyme-Linked Immuno Sorbent Assay). The step of adding a solution containing a candidate protein to a solid-phased target nucleic acid

can be specifically carried out by flowing a solution containing the objective binding protein on the target nucleic acid molecule immobilized on a plate. Immobilization of the target nucleic acid molecule can be achieved by using various existing immobilization methods, such as by providing a nucleic acid probe containing a biotin-modified target nucleic acid molecule to a streptavidin-coated well plate. For detailed conditions of the experiments, the experiment methods described in detail in the section of Examples in this description can be referred to. In RPB-ELISA, a value obtained by subtracting background signal (luminescence signal value obtained with an objective PPR protein without adding the target RNA) from luminescence obtained with a sample containing the objective PPR protein and the target RNA thereof can be used as the binding power of the objective PPR protein and the target RNA thereof.

[0054]

[Use of PPR protein]

(Complex and fusion protein)

The PPR motif or PPR protein provided by the present invention can be made into a complex by binding a functional region. The PPR motif or PPR protein can also be linked with a proteinaceous functional region to form a fusion protein. The functional region refers to a part having such a function as a specific biological function exerted in a living body or cell, for example, enzymatic function, catalytic function, inhibitory function, promotion function, etc., or a function as a marker. Such a region consists of, for example, a protein, peptide, nucleic acid, physiologically active substance, or drug. In the following explanations, the complex of the present invention may be explained with reference to a fusion protein as an example, but those skilled in the art may also understand complexes other than fusion protein according to the explanations.

[0055]

In one of the preferred embodiments, the functional region is a ribonuclease (RNase). Examples of RNase are RNase A (e.g., bovine pancreatic ribonuclease A, PDB 2AAS), and RNase H.

[0056]

In one of the preferred embodiments, the functional region is a fluorescent protein. Examples of fluorescent protein are mCherry, EGFP, GFP, Sirius, EBFP, ECFP, mTurquoise, TagCFP, AmCyan, mTFP1, MidoriishiCyan, CFP, TurboGFP, AcGFP, TagGFP, Azami-Green, ZsGreen, EmGFP, HyPer, TagYFP, EYFP, Venus, YFP, PhiYFP, PhiYFP-m, TurboYFP, ZsYellow, mBanana, KusabiraOrange, mOrange, TurboRFP, DsRed-Express, DsRed2, TagRFP, DsRed-Monomer, AsRed2, mStrawberry,

TurboFP602, mRFP1, JRed, KillerRed, HcRed, KeimaRed, mRaspberry, mPlum, PS-CFP, Dendra2, Kaede, EosFP, and KikumeGR. A preferred example is mClover3 in view of improvement of aggregation and/or efficient localization to the nuclei as a fusion protein.

[0057]

In one preferred embodiment, when the target is mRNA, the functional region is a functional domain that enhances expression amount of a protein from the target mRNA (WO2017/209122). The functional domain that enhances expression amount of a protein from mRNA may be, for example, all or a functional part of a functional domain of a protein known to directly or indirectly promote translation of mRNA. More specifically, it may be a domain that directs ribosomes to mRNA, domain associated with initiating or promoting translation of mRNA, domain associated with transporting mRNA out of the nucleus, domain associated with binding to the endoplasmic reticulum membrane, domain containing an endoplasmic reticulum (ER) retention signal sequence, or domain containing an endoplasmic reticulum signal sequence. More specifically, the domain that directs ribosomes to mRNA mentioned above may be a domain comprising all or a functional part of a polypeptide selected from the group consisting of density-regulated protein (DENR), malignant T-cell amplified sequence 1 (MCT-1), transcriptionally-controlled tumor protein (TPT1), and Lerepo4 (zinc finger CCHH-domain). The domain associated with translation initiation or translation promotion of mRNA mentioned above may be a domain comprising all or a functional part of a polypeptide selected from the group consisting of eIF4E and eIF4G. The domain associated with transporting mRNA out of the nucleus mentioned above may be a domain containing all or a functional part of stem-loop binding protein (SLBP). The domain associated with binding to the endoplasmic reticulum membrane mentioned above may be a domain comprising all or a functional part of a polypeptide selected from the group consisting of SEC61B, translocation associated protein alpha (TRAP-alpha), SR-alpha, Dia1 (cytochrome b5 reductase 3), and p180. The endoplasmic reticulum retention signal (ER retention signal) sequence mentioned above may be a signal sequence comprising the KDEL (SEQ ID NO: 55) or KEEL (SEQ ID NO: 56) sequence. The endoplasmic reticulum signal sequence mentioned above may be a signal sequence including MGWSCILFLVATATGAHS (SEQ ID NO: 57).

[0058]

In the present invention, the functional region may be fused to the PPR protein on the N-terminal side or the C-terminal side, or on both the N-terminal side and the

C-terminal side. The complex or fusion protein may include a plurality of functional regions (e.g., 2 to 5). Further, the complex or fusion protein according to the present invention may consist of the functional region and PPR protein indirectly fused via a linker or the like.

[0059]

(Nucleic acid encoding PPR protein etc., vector, and cell)

The present invention also provides a nucleic acid encoding the PPR motif, PPR protein or fusion protein mentioned above, and a vector containing such a nucleic acid (e.g., vector for amplification, and expression vector). As the host of the vector for amplification, *E. coli* or yeast may be used. In this description, expression vector means a vector containing, for example, a DNA having a promoter sequence, DNA encoding a desired protein, and DNA having a terminator sequence from the upstream side, but they need not necessarily be arranged in this order, so long as the desired function is exerted. In the present invention, recombinant vectors prepared by using various vectors that may be normally used by those skilled in the art may be used.

[0060]

The PPR protein or fusion protein of the present invention can function in eukaryotic (e.g., animal, plant, microbe (yeast, etc.), and protozoan) cells. The fusion protein of the present invention can function, in particular, in animal cells (in vitro or in vivo). Examples of animal cells into which the PPR protein or fusion protein of the present invention, or a vector expressing it can be introduced include, for example, cells derived from humans, monkeys, pigs, cows, horses, dogs, cats, mice, and rats. Examples of cultured cells into which the PPR protein or fusion protein of the present invention or a vector expressing it can be introduced include, for example, Chinese hamster ovary (CHO) cells, COS-1 cells, COS-7 cells, VERO (ATCC CCL-81) cells, BHK cells, canine kidney-derived MDCK cells, hamster AV-12-664 cells, HeLa cells, WI38 cells, 293 cells, 293T cells, and PER.C6 cells, but not limited to these.

[0061]

(Use)

With the PPR protein or fusion protein of the present invention, a functional region may be delivered to the inside of a living body or cells and made to function in a nucleic acid sequence-specific manner. A complex linked with a marker such as GFP may be used to visualize a desired RNA in a living body.

[0062]

With the PPR protein or fusion protein of the present invention, a nucleic acid can be modified or disrupted in a nucleic acid sequence-specific manner in the inside

of cells or living bodies, and a new function may be conferred. In particular, RNA-binding PPR proteins are involved in all the RNA processing steps found in the organelles, such as cleavage, RNA edition, translation, splicing, and RNA stabilization. Accordingly, such uses of the method concerning modification of PPR proteins provided by the present invention, as well as the PPR motif and PPR protein provided by the present invention as mentioned below can be expected in a variety of fields.

[0063]

(1) Medical care

- Creation of a PPR protein that recognizes and binds to a specific RNA associated with a specific disease. Analysis of a target sequence and associated proteins for a specific RNA. The results of the analysis can be used to identify compounds for the treatment of the disease.

[0064]

For example, it is known that, in animals, abnormalities in the PPR protein identified as LRPPRC cause Leigh syndrome, French Canadian type (LSFC, Leigh syndrome, subacute necrotizing encephalomyelopathy). The present invention may contribute to the treatment (prevention, therapeutic treatment, or inhibition of progression) of LSFC. Many of the existing PPR proteins work to specify edition sites for RNA manipulation (conversion of genetic information on RNA, often C to U). The PPR proteins of this type have an additional motif that is suggested to interact with RNA editing enzymes on the C-terminal side. PPR proteins having this structure are expected to enable introduction of base polymorphism or treatment of a disease or condition caused by base polymorphism.

[0065]

- Creation of cells with controlled RNA repression/expression. Such cells include stem cells of which differentiation or undifferentiation state is monitored (e.g., iPS cells), model cells for evaluation of cosmetics, and cells in which the expression of functional RNA can be turned on or off for the purpose of elucidating action mechanism and pharmacological testing for drug discovery.

[0066]

- Preparation of a PPR protein that specifically binds to a specific RNA associated with a particular disease. Such a PPR protein is introduced into a cell using a plasmid, virus vector, mRNA, or purified protein, and an RNA function that causes a disease can be changed (improved) by binding of the PPR protein to the target RNA in the cell. Examples of the mechanism of changing the function include, for example, change of the RNA structure by binding, knockdown by decomposition,

change of the splicing reaction by splicing, base substitution, and so forth.

[0067]

## (2) Agriculture, forestry and fishery

- Improvement of yield and quality of crops, forest products and marine products.

- Breeding of organisms with improved disease resistance, improved environmental tolerance, or improved or new function.

[0068]

For example, concerning hybrid firstgeneration (F1) plant crops, an F1 plant may be artificially created by using stabilization of mitochondrial RNA and translation control by PPR proteins so that yield and quality of the crops may be improved. RNA manipulation and genome edition using PPR proteins more accurately and quickly enable variety improvement and breeding (genetic improvement of organisms) of organisms compared with conventional techniques. In addition, it can be said that RNA manipulation and genome editing using PPR proteins are similar to the classical breeding methods such as selection of mutants and backcrossing, since they do not transform traits with a foreign gene as in genetic recombination, but they are techniques using RNA and genomes originally possessed by plants and animals. Therefore, they can also surely and quickly cope with global-scale food and environmental problems.

[0069]

## (3) Chemistry

- Control of protein expression amount by manipulating DNA and RNA in the production of useful substances using microorganisms, cultured cells, plant bodies, and animal bodies (e.g., insect bodies). Productivity of useful substances can be thereby improved. Examples of the useful substances are proteinaceous substances such as antibodies, vaccines, and enzymes, as well as relatively low-molecular weight compounds such as pharmaceutical intermediates, fragrances, and dyes.

[0070]

- Improvement of production efficiency of biofuel by modification of metabolic pathways of algae and microorganisms.

## Examples

[0071]

[Example 1: Intracellular analysis of fluorescent protein-fused PPR proteins]

(Design of motifs)

The target sequence was CAGCAGCAGCAGCAGCAG (SEQ ID NO: 1) including six repeats of CAG sequence. The base recognized by a PPR motif is determined by the 1st, 4th and ii-th amino acids in the sequence. The PPR motif that recognizes cytosine contained 1st is valine, 4th asparagine, and ii-th serine, the PPR motif that recognizes adenine contained 1st is valine, 4th threonine, and ii-th asparagine, and the PPR motif that recognizes guanine contained 1st is valine, 4th threonine, and ii-th aspartic acid, respectively. For the PPR motif that recognizes uracil, 1st valine, 4th asparagine, and ii-th aspartic acid can be used.

[0072]

Further, as the 6th and 9th amino acids of the 1st motif that recognizes cytosine (Mutated motif in Fig. 1A), there were selected leucine and glycine (C\_6L9G, PPRcag\_1, SEQ ID NO: 2 mentioned above) as a typical combination, and leucine and glutamic acid (C\_6L9E, PPRcag\_2, SEQ ID NO: 3), asparagine and glutamine (C\_6N9Q, PPRcag\_3, SEQ ID NO: 4), asparagine and glutamic acid (C\_6N9E, PPRcag\_4, SEQ ID NO: 5), asparagine and lysine (C\_6N9K, PPRcag\_5, SEQ ID NO: 6), and aspartic acid and glycine (C\_6D9G, PPRcag\_6, SEQ ID NO: 7) were selected as mutant types (Fig. 1B). These PPR motif sequences were arranged so that the encoded proteins bind to the CAGCAGCAGCAGCAGCAG sequence (SEQ ID NO: 1 mentioned above), and PPR genes (SEQ ID NOS: 11 to 16) were prepared. In order to efficiently and accurately ligate the 18 of DNAs encoding the respective PPR motifs, the highly frequently found amino acids were selected as the amino acids other than the 1st, 4th, 6th, 9th, and ii-th amino acids in the PPR motifs for cytosine, adenine, and guanine, respectively, as described above (SEQ ID NOS: 8 and 9, see Patent document 1 mentioned above).

[0073]

(Preparation of plasmids)

Plasmids containing each of the PPR genes were constructed by using the Golden Gate method. In more detail, 10 kinds of intermediate vectors Dest-a, b, c, d, e, f, g, h, i, and j, designed to be seamlessly ligated in sequence, were prepared, and 20 kinds of motifs consisting of one motif or two motifs (single PPR motifs corresponding to A, C, G, and U, and double PPR motifs corresponding to AA, AC, AG, AU, CA, CC, CG, CU, GA, GC, GG, GU, UA, UC, UG, and UU) were inserted into 10 kinds of the vectors to produce 200 kinds of parts.

[0074]

There were prepared

gaagacataaactccgtggtcacATACagagaccaaggtctcaGTGGtcacatacatgtcttc (SEQ ID NO:

43) as Dest-a,

gaagacatATACagagaccaaggtctcaGTGGtgacataatgtcttc (SEQ ID NO: 44) as Dest-b,  
gaagacatcATACagagaccaaggtctcaGTGGttacatatgtcttc (SEQ ID NO: 45) as Dest-c,  
gaagacatacATACagagaccaaggtctcaGTGGttacaatgtcttc (SEQ ID NO: 46) as Dest-d,  
gaagacattacATACagagaccaaggtctcaGTGGtgacatgtcttc (SEQ ID NO: 47) as Dest-e,  
gaagacattgacATACagagaccaaggtctcaGTGGttaatgtcttc (SEQ ID NO: 48) as Dest-f,  
gaagacatgttacATACagagaccaaggtctcaGTGGtcatgtcttc (SEQ ID NO: 49) as Dest-g,  
gaagacatggtcacATACagagaccaaggtctcaGTGGtatgtcttc (SEQ ID NO: 50) as Dest-h,  
gaagacattggttacATACagagaccaaggtctcaGTGGatgtcttc (SEQ ID NO: 51) as Dest-i, and  
gaagacatgtggtgacATACagagaccaaggtctcaGTGGtcttc (SEQ ID NO: 52) as Dest-j  
by a gene synthesis technique, and cloned into pUC57-kan.

[0075]

Dest-a to Dest-j were selected according to the target sequence, and cloned into the vector by the Golden Gate reaction. The vector used here was designed so that the amino acid sequence of MGNSV (SEQ ID NO: 53) was added to the N-terminus of the 18 PPR sequences linked together, and ELTYNTLISGLGKAGRARDPPV (SEQ ID NO: 54) to the C-terminus of the same. It was confirmed that the correct size genes were cloned, and the sequences of the cloned genes were confirmed by sequencing.

[0076]

(Detection of expression in cells)

The expression plasmid pcDNA3.1 for expression in cultured animal cells contains the CMV promoter and SV40 poly-A signal sequence, and a gene to be expressed can be inserted between them. To detect the expression of PPR proteins in cells, PPR proteins fused with a fluorescent protein were expressed, and aggregation and nuclear localization thereof in the cells were analyzed on the basis of fluorescent images thereof. Fused protein genes comprising those for EGFP, nuclear localization signal sequence, PPR protein, and FLAG epitope tag fused together in this order from the N-terminus side were inserted into pcDNA3.1 (SEQ ID NOS: 17 to 22). Protein genes comprising those for mClover3, PPR protein, nuclear localization signal sequence, and FLAG epitope tag fused together in this order from the N-terminus side were also inserted into pcDNA3.1 (SEQ ID NOS: 23 to 28). Plasmids not containing PPR were also prepared as control (SEQ ID NOS: 35 and 36).

[0077]

The HEK293T cells were inoculated at a density of  $1 \times 10^6$  cells in 10 cm dish containing 9 mL of DMEM, and 1 mL of FBS, and cultured in an environment of 37°C

and 5% CO<sub>2</sub> for 2 days, and then the cells were collected. The collected cells were inoculated on a PLL-coated 96-well plate at a density of 4 x 10<sup>4</sup> cells/well, and cultured in an environment of 37°C and 5% CO<sub>2</sub> for 1 day. A mixture of 200 ng of the plasmid DNA, 0.6 µL of Fugene (registered trademark)-HD (Promega, E2311), and 200 µL of Opti-MEM was prepared, the whole volume thereof was added to each well, and culture was performed in an environment of 37°C and 5% CO<sub>2</sub> for 1 day. After the culture, the medium was removed, each well was washed once with 50 µL of PBS, then 1 µL of Hoechst (1 mg/mL, Dojin Chemical, 346-07951) and 50 µL of PBS were added to each well, the plate was left under an environment of 37°C and 5% CO<sub>2</sub> for 10 minutes, and then each well was washed with 50 µL of PBS. After the washing, 50 µL of PBS was added, and GFP fluorescence and Hoechst fluorescence images of each well were obtained by using a DMI8 fluorescence microscope (Leica).

[0078]

The results are shown in Fig. 2. Expression in the cells of the PPR proteins fused with EGFP and nuclear localization signal sequence was confirmed, and as a result, it was confirmed that PPRcag\_1 (6L9G) and PPRcag\_2 (6L9E) did not localize to the nuclei, but strongly aggregated around the nuclei. On the other hand, PPRcag\_3 (6N9Q), PPRcag\_4 (6N9E), PPRcag\_5 (6N9K), and PPRcag\_6 (6D9G) did not localize to the nuclei, although their aggregation was weak. When mClover3 was fused, it was confirmed that PPRcag\_1 (6L9G) and PPRcag\_2 (6L9E) localized to the nuclei, but aggregated in the nuclei. In contrast, PPRcag\_3 (6N9Q), PPRcag\_4 (6N9E), PPRcag\_5 (6N9K), and PPRcag\_6 (6D9G) localized to the nuclei, and did not aggregate. Therefore, it was found that 6N9E, 6N9Q, 6N9K, and 6D9G mutations are favorable for improving aggregation, and mClover3 is better than EGFP for efficient localization to the nuclei.

[0079]

[Example 2: RNA binding analysis of CAG-binding PPR proteins]

To confirm binding of PPRcag\_1, PPRcag\_2, PPRcag\_3, PPRcag\_4, PPRcag\_5, and PPRcag\_6 to their target RNAs, recombinant proteins were prepared, and binding experiments were performed.

[0080]

Protein genes were designed for the respective PPR proteins fused with luciferase on the N-terminus side and 6 x histidine tag sequence on the C-terminus side, and cloned into an *E. coli* expression plasmid (SEQ ID NOS: 29 to 34). Nluc-Hisx6 protein gene not containing PPR protein was also prepared as a control (SEQ ID NOS: 37).

[0081]

The *E. coli* Rosetta (DE3) strain was transformed with the completed plasmids. The *E. coli* was cultured in 2 mL of the LB medium containing 100 µg/mL ampicillin at 37°C for 12 hours. When OD<sub>600</sub> reached 0.5 to 0.8, the culture medium was transferred to an incubator at 15°C, and left standing for 30 minutes. Then, 100 µL of an IPTG solution was added (IPTG final concentration, 0.1 mM), and the culture was further continued at 15°C for 16 hours. An *E. coli* pellet was collected by centrifugation at 5,000 x g and 4°C for 10 minutes, 1.5 mL of a lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% NP-40, 1 mM MgCl<sub>2</sub>, 2 mg/mL lysozyme, 1 mM PMSF, 2 µL of DNase) was added to the pellet, and the mixture was frozen at -80°C for 20 minutes. The cells were cryodisrupted with permeabilization at 25°C for 30 minutes. The disrupted cell mixture was then centrifuged at 3,700 rpm and 4°C for 15 minutes, and the supernatant containing soluble PPR protein (*E. coli* lysate) was collected.

[0082]

The binding experiment for PPR protein and RNA was performed by an experimental method for binding of PPR protein and biotinylated RNA on a streptavidin plate.

[0083]

There were synthesized RNA probes of 30-base RNAs containing the target CAGx6 sequence, untargeted CGGx6, CUGx6, and CCGx6, and the D1b (UGGUGUAUCUUGUCUUUA) sequence (positions 8 to 25 of SEQ ID NO: 42) modified with biotin at the 5'-end (in that order, SEQ ID NOS: 38 to 42, respectively) (Grainer). To a streptavidin-coated plate (Thermo Fisher, Cat. No. 15502), the 5'-end biotinylated RNA probes were added, reaction was allowed at room temperature for 30 minutes, and the plate was washed with a probe washing buffer (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5% NP-40, 1 mM DTT, 0.1% BSA). For background measurement, wells to which biotinylated RNA was not added, but the lysis buffer was added were also prepared. Then, a blocking buffer (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5% NP-40, 1 mM DTT, 1% BSA) was added, and the plate surface was blocked at room temperature for 30 minutes. Then, 100 µL of the *E. coli* lysate containing luciferase-fused PPR protein having a luminescence of  $1.5 \times 10^8$  LU/µL was added, and the binding reaction was allowed at room temperature for 30 minutes. Then, the plate was washed 5 times with 200 µL of a washing buffer (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.5% NP-40, 1 mM DTT). To each well, 40 µL of luciferase substrate (Promega, E151A) diluted 2,500-fold with

the washing buffer was added, reaction was allowed for 5 minutes, and then luminescence was measured with a plate reader (PerkinElmer, Cat. No. 5103-35).

[0084]

The results are shown in Fig. 3. It was found that all PPRs specifically bind to the target, CAGx6. The binding powers of PPRcag\_2, PPRcag\_3, PPRcag\_4, PPRcag\_5, and PPRcag\_6 to the target sequence were substantially equivalent, about 80%, about 60%, about 120%, and about 130%, respectively, compared with that of PPRcag\_1. These results indicate that the binding performance was not substantially changed by the mutations except for PPRcag\_4.

[0085]

[Example 3: Control of aggregation of PPR protein]

A PPR protein using V2 motif (SEQ ID NO: 61 for nucleotide sequence, and SEQ ID NO: 62 for amino acid sequence) and a PPR protein using v3.2 motif (SEQ ID NO: 63 for nucleotide sequence, and SEQ ID NO: 64 for amino acid sequence) were prepared in an *E. coli* expression system, respectively, purified, and separated by gel filtration chromatography. The v2 motif refers to the PPR motifs having the sequence of SEQ ID NO: 2, and SEQ ID NOS: 8 to 10, and the v3.2 motif refers to the PPR motifs having the sequence of SEQ ID NO: 4 and SEQ ID NOS: 58 to 60 in the case of the first motif from the N-terminus side, or in the other cases, the PPR motif for adenine comprising the sequence of SEQ ID NO: 8 having a substitution of 15th aspartic acid with lysine, or the PPR motifs for bases other than adenine comprising a sequence selected from SEQ ID NOS: 2, 9, and 10.

[0086]

(Expression and purification of proteins)

The *E. coli* Rosetta strain was transformed with pE-SUMOpro Kan plasmid containing a DNA sequence encoding the objective PPR protein, and cultured at 37°C, then the temperature was lowered to 20°C when OD<sub>600</sub> reached 0.6, and IPTG was added at a final concentration of 0.5 mM so that the objective PPR was expressed in the *E. coli* cells as SUMO-fused protein. The cells were cultured overnight, then collected by centrifugation, and resuspended in a lysis buffer (50mM Tris-HCl, pH 8.0, 500mM NaCl). The *E. coli* cells were disrupted by sonication, and centrifuged at 17,000g for 30 minutes, then the supernatant fraction was applied to an Ni-Agarose column, the column was washed with the lysis buffer containing 20 mM imidazole, and then the SUMO-fused objective PPR protein was eluted with the lysis buffer containing 400 mM imidazole. After the elution, the SUMO protein was cleaved from the objective PPR protein with Ulp1, and at the same time, the protein solution

was substituted with an ion-exchange buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl) by dialysis. Subsequently, cation exchange chromatography was performed by using SP column. After application to the column, proteins were eluted with gradually increasing NaCl concentration of from 200 mM to 1 M. The fraction containing the objective PPR protein was subjected to final purification by gel filtration chromatography using Superdex 200 column. The objective PPR protein eluted from the ion exchange column was applied to the gel filtration column equilibrated with a gel filtration buffer (25 mM HEPES, pH 7.5, 200 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)). Finally, the fraction containing the objective PPR protein was concentrated, frozen in liquid nitrogen, and stored at -80°C until used for the next analysis.

[0087]

(Gel filtration chromatography)

The purified recombinant PPR protein was prepared at a concentration of 1 mg/ml. For gel filtration chromatography, Superdex 200 increase 10/300 GL (GE Healthcare) was used. To the gel filtration column equilibrated with 25 mM HEPES pH7.5, 200 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), the prepared protein was applied, and the absorbance of the solution eluted from the gel filtration column was measured at 280 nm to analyze the properties of the protein.

[0088]

(Results)

The results are shown in Fig. 4. The smaller volume of the elution fraction (Elution vol.) means a larger molecular size. The proteins using v2 were eluted in 8 to 10 mL of elution fractions, whereas the peaks of the proteins using v3.2 were observed in elution fractions of 12 to 14 mL. This result suggested possibility that the proteins using v2 aggregated due to the larger protein size thereof, and the aggregation was improved in the proteins using v3.2.

[0089]

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

[0090]

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that the prior art forms part of the common general knowledge in Australia.

## Claims

[Claim 1]

A polypeptide, which is any one of the following polypeptides:

(C-1) a polypeptide consisting of any one of the sequences of SEQ ID NOs: 4 to 7;

(C-2) a polypeptide consisting of any one of the sequences of SEQ ID NOs: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, having a cytosine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(C-3) a polypeptide having a sequence identity of at least 70% to any one of the sequences of SEQ ID NOs: 4 to 7, provided that the amino acids at positions 1, 4, 6, and 34 are identical, having a cytosine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(A-1) a polypeptide consisting of the sequence of SEQ ID NO: 8 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(A-2) a polypeptide consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, having an adenine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(A-3) a polypeptide having a sequence identity of at least 70% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, having an adenine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(G-1) a polypeptide consisting of the sequence of SEQ ID NO: 9 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(G-2) a polypeptide consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, having a guanine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(G-3) a polypeptide having a sequence identity of at least 70% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, having a guanine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(U-1) a polypeptide consisting of the sequence of SEQ ID NO: 10 having a substitution of the amino acid at position 6 with asparagine or aspartic acid;

(U-2) a polypeptide consisting of the sequence of (U-1) having a substitution, deletion,

or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, and 34, having a uracil-binding property, and being a PPR motif represented by the formula 1 mentioned below; and

(U-3) a polypeptide having a sequence identity of at least 70% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, and 34 are identical, having a uracil-binding property, and being a PPR motif represented by the formula 1 mentioned below.

**(Helix A)-X-(Helix B)-L** (Formula 1)

wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

**A<sub>1</sub>-A<sub>2</sub>-A<sub>3</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>-A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>** (Formula 2)

wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

**L<sub>vii</sub>-L<sub>vi</sub>-L<sub>v</sub>-L<sub>iv</sub>-L<sub>iii</sub>-L<sub>ii</sub>-L<sub>i</sub>** (Formula 3)

wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist.

[Claim 2]

A polypeptide, which is any one of the following PPR motifs:

(C-1) a polypeptide consisting of any one of the sequences of SEQ ID NOs: 4 to 7;

(C-2) a polypeptide consisting of any one of the sequences of SEQ ID NOs: 4 to 7 having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, having a cytosine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(C-3) a polypeptide having a sequence identity of at least 70% to any one of the sequences of SEQ ID NOs: 4 to 7, provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, having a cytosine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(A-1) a polypeptide consisting of the sequence of SEQ ID NO: 8 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(A-2) a polypeptide consisting of the sequence of (A-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, having an adenine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(A-3) a polypeptide having a sequence identity of at least 70% to the sequence of (A-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, having an adenine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(G-1) a polypeptide consisting of the sequence of SEQ ID NO: 9 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(G-2) a polypeptide consisting of the sequence of (G-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, having a guanine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(G-3) a polypeptide having a sequence identity of at least 70% to the sequence of (G-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, having a guanine-binding property, and being a PPR motif represented by the formula 1 mentioned below;

(U-1) a polypeptide consisting of the sequence of SEQ ID NO: 10 having such substitutions of the amino acids at positions 6 and 9 that any one of the combinations defined below is satisfied;

(U-2) a polypeptide consisting of the sequence of (U-1) having a substitution, deletion, or addition of 1 to 9 amino acids other than the amino acids at positions 1, 4, 6, 9, and 34, having a uracil-binding property, and being a PPR motif represented by the formula 1 mentioned below; and

(U-3) a polypeptide having a sequence identity of at least 70% to the sequence of (U-1), provided that the amino acids at positions 1, 4, 6, 9, and 34 are identical, having a uracil-binding property, and being a PPR motif represented by the formula 1 mentioned below.

- a combination of asparagine as the amino acid at position 6 and glutamic acid as the amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and glutamine as the

amino acid at position 9,

- a combination of asparagine as the amino acid at position 6 and lysine as the amino acid at position 9, and

- a combination of aspartic acid as the amino acid at position 6 and glycine as the amino acid at position 9.

**(Helix A)-X-(Helix B)-L** (Formula 1)

wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

**A<sub>1</sub>-A<sub>2</sub>-A<sub>3</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>-A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>** (Formula 2)

wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

**L<sub>vii</sub>-L<sub>vi</sub>-L<sub>v</sub>-L<sub>iv</sub>-L<sub>iii</sub>-L<sub>ii</sub>-L<sub>i</sub>** (Formula 3)

wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist.

[Claim 3]

The polypeptide according to claim 1 or 2, which is any one of the following PPR motifs:

(C-4) a polypeptide consisting of the sequence of SEQ ID NO: 4;

(A-4) a polypeptide consisting of the sequence of SEQ ID NO: 58;

(G-4) a polypeptide consisting of the sequence of SEQ ID NO: 59; and

(U-4) a polypeptide consisting of the sequence of SEQ ID NO: 60.

[Claim 4]

Use of the polypeptide according to any one of claims 1 to 3 in a PPR protein as the first PPR motif from the N-terminus.

[Claim 5]

The use according to claim 4, which is for reducing aggregation of the PPR protein.

[Claim 6]

A protein capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises 1 to 30 of PPR motifs represented by the formula 1 mentioned below, and

wherein the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus is a hydrophilic amino acid; or

the polypeptide according to any one of claims 1 and 3:

**(Helix A)-X-(Helix B)-L** (Formula 1)

wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

**A<sub>1</sub>-A<sub>2</sub>-A<sub>3</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>-A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>** (Formula 2)

wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;

**L<sub>vii</sub>-L<sub>vi</sub>-L<sub>v</sub>-L<sub>iv</sub>-L<sub>iii</sub>-L<sub>ii</sub>-L<sub>i</sub>** (Formula 3)

wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist,

wherein, the combination of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> amino acids in M<sub>1</sub> is selected from the following:

(2-1) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, glycine, and aspartic acid in this order;

(2-2) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of glutamic acid, glycine, and aspartic acid in this order;

(2-3) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, glycine, and asparagine in this order;

(2-4) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of glutamic

acid, glycine, and asparagine in this order;

(2-5) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, glycine, and serine in this order;

(2-6) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, isoleucine, and an arbitrary amino acid in this order;

(2-7) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, isoleucine, and asparagine in this order;

(2-8) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and an arbitrary amino acid in this order;

(2-9) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and aspartic acid in this order;

(2-10) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, leucine, and lysine in this order;

(2-11) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, methionine, and an arbitrary amino acid in this order;

(2-12) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, methionine, and aspartic acid in this order;

(2-13) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, methionine, and aspartic acid in this order;

(2-14) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and an arbitrary amino acid in this order;

(2-15) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and aspartic acid in this order;

(2-16) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, asparagine, and aspartic acid in this order;

(2-17) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glycine, asparagine, and aspartic acid in this order;

(2-18) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and aspartic acid in this order;

(2-19) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of threonine, asparagine, and aspartic acid in this order;

(2-20) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and aspartic acid in this order;

(2-21) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, asparagine, and aspartic acid in this order;

(2-22) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an

arbitrary amino acid, asparagine, and asparagine in this order;

(2-23) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and asparagine in this order;

(2-24) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of serine, asparagine, and asparagine in this order;

(2-25) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and asparagine in this order;

(2-26) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and serine in this order;

(2-27) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and serine in this order;

(2-28) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and threonine in this order;

(2-29) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and threonine in this order;

(2-30) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and tryptophan in this order;

(2-31) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and tryptophan in this order;

(2-32) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and an arbitrary amino acid in this order;

(2-33) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and aspartic acid in this order;

(2-34) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, proline, and aspartic acid in this order;

(2-35) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, proline, and aspartic acid in this order;

(2-36) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and an arbitrary amino acid in this order;

(2-37) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and asparagine in this order;

(2-38) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, serine, and asparagine in this order;

(2-39) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, serine, and asparagine in this order;

(2-40) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an

- arbitrary amino acid, threonine, and an arbitrary amino acid in this order;
- (2-41) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, threonine, and aspartic acid in this order;
- (2-42) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of valine, threonine, and aspartic acid in this order;
- (2-43) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, threonine, and asparagine in this order;
- (2-44) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of phenylalanine, threonine, and asparagine in this order;
- (2-45) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of isoleucine, threonine, and asparagine in this order;
- (2-46) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of valine, threonine, and asparagine in this order;
- (2-47) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, valine, and an arbitrary amino acid in this order;
- (2-48) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of isoleucine, valine, and aspartic acid in this order;
- (2-49) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, valine, and glycine in this order;
- (2-50) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, valine, and threonine in this order;
- (3-7) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of lysine, threonine, and aspartic acid in this order;
- (3-12) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of threonine, threonine, and asparagine in this order; or
- (3-16) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of leucine, threonine, and aspartic acid in this order.

[Claim 7]

The protein according to claim 6, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is a hydrophilic amino acid or glycine.

[Claim 8]

The protein according to claim 6 or 7, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> is asparagine or aspartic acid.

[Claim 9]

The protein according to any one of claims 6 to 8, wherein the A<sub>9</sub> amino acid of M<sub>1</sub> is glutamine, glutamic acid, lysine, or glycine.

[Claim 10]

The protein according to any one of claims 6 to 9, wherein the A<sub>6</sub> amino acid of M<sub>1</sub> and the A<sub>9</sub> amino acid of M<sub>1</sub> correspond to any of the following combinations:

- combination of asparagine as the A<sub>6</sub> amino acid and glutamic acid as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and glutamine as the A<sub>9</sub> amino acid,
- combination of asparagine as the A<sub>6</sub> amino acid and lysine as the A<sub>9</sub> amino acid, and
- combination of aspartic acid as the A<sub>6</sub> amino acid and glycine as the A<sub>9</sub> amino acid.

[Claim 11]

A fusion protein of  
 at least one selected from the group consisting of a fluorescent protein, a nuclear localization signal peptide, and a tag protein, and  
 a PPR protein containing the polypeptide according to any one of claims 1 to 3 as the first PPR motif from the N-terminus, or the protein according to any one of claims 6 to 10.

[Claim 12]

A method for reducing aggregation of a PPR protein containing the PPR motif represented by the formula 1 mentioned below and capable of binding to a target nucleic acid having a specific nucleotide sequence, which comprises making the A<sub>6</sub> amino acid of the first PPR motif (M<sub>1</sub>) from the N-terminus a hydrophilic amino acid:

**(Helix A)-X-(Helix B)-L** (Formula 1)

wherein, in the formula:

Helix A is a moiety of 12-amino acid length capable of forming an  $\alpha$ -helix structure, and is represented by the formula 2;

**A<sub>1</sub>-A<sub>2</sub>-A<sub>3</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>-A<sub>10</sub>-A<sub>11</sub>-A<sub>12</sub>** (Formula 2)

wherein, in the formula 2, A<sub>1</sub> to A<sub>12</sub> independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an  $\alpha$ -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L<sub>iii</sub> to L<sub>vii</sub> may not exist,

wherein, the combination of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> amino acids in M<sub>1</sub> is selected from the following:

(2-1) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, glycine, and aspartic acid in this order;

(2-2) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of glutamic acid, glycine, and aspartic acid in this order;

(2-3) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, glycine, and asparagine in this order;

(2-4) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of glutamic acid, glycine, and asparagine in this order;

(2-5) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, glycine, and serine in this order;

(2-6) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, isoleucine, and an arbitrary amino acid in this order;

(2-7) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, isoleucine, and asparagine in this order;

(2-8) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, leucine, and an arbitrary amino acid in this order;

(2-9) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, leucine, and aspartic acid in this order;

(2-10) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, leucine, and lysine in this order;

(2-11) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, methionine, and an arbitrary amino acid in this order;

(2-12) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, methionine, and aspartic acid in this order;

- (2-13) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, methionine, and aspartic acid in this order;
- (2-14) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and an arbitrary amino acid in this order;
- (2-15) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and aspartic acid in this order;
- (2-16) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, asparagine, and aspartic acid in this order;
- (2-17) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of glycine, asparagine, and aspartic acid in this order;
- (2-18) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and aspartic acid in this order;
- (2-19) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of threonine, asparagine, and aspartic acid in this order;
- (2-20) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and aspartic acid in this order;
- (2-21) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, asparagine, and aspartic acid in this order;
- (2-22) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and asparagine in this order;
- (2-23) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and asparagine in this order;
- (2-24) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of serine, asparagine, and asparagine in this order;
- (2-25) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and asparagine in this order;
- (2-26) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and serine in this order;
- (2-27) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and serine in this order;
- (2-28) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and threonine in this order;
- (2-29) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, asparagine, and threonine in this order;
- (2-30) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, asparagine, and tryptophan in this order;

- (2-31) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, asparagine, and tryptophan in this order;
- (2-32) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and an arbitrary amino acid in this order;
- (2-33) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, proline, and aspartic acid in this order;
- (2-34) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, proline, and aspartic acid in this order;
- (2-35) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of tyrosine, proline, and aspartic acid in this order;
- (2-36) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and an arbitrary amino acid in this order;
- (2-37) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, serine, and asparagine in this order;
- (2-38) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, serine, and asparagine in this order;
- (2-39) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, serine, and asparagine in this order;
- (2-40) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and an arbitrary amino acid in this order;
- (2-41) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and aspartic acid in this order;
- (2-42) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and aspartic acid in this order;
- (2-43) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, threonine, and asparagine in this order;
- (2-44) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of phenylalanine, threonine, and asparagine in this order;
- (2-45) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, threonine, and asparagine in this order;
- (2-46) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of valine, threonine, and asparagine in this order;
- (2-47) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of an arbitrary amino acid, valine, and an arbitrary amino acid in this order;
- (2-48) the combination of the three amino acids of  $A_1$ ,  $A_4$ , and  $L_{ii}$  consists of isoleucine, valine, and aspartic acid in this order;

- (2-49) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, valine, and glycine in this order;
- (2-50) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of an arbitrary amino acid, valine, and threonine in this order;
- (3-7) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of lysine, threonine, and aspartic acid in this order;
- (3-12) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of threonine, threonine, and asparagine in this order; or
- (3-16) the combination of the three amino acids of A<sub>1</sub>, A<sub>4</sub>, and L<sub>ii</sub> consists of leucine, threonine, and aspartic acid in this order.

[Claim 13]

A method for detecting a nucleic acid, which uses the protein according to any one of claims 6 to 10, or the fusion protein according to claim 11.

[Claim 14]

A nucleic acid encoding the polypeptide according to any one of claims 1 to 3, or the protein according to any one of claims 6 to 10.

[Claim 15]

A vector comprising the nucleic acid according to claim 14.

[Claim 16]

A cell (except for human individual) containing the vector according to claim 15.

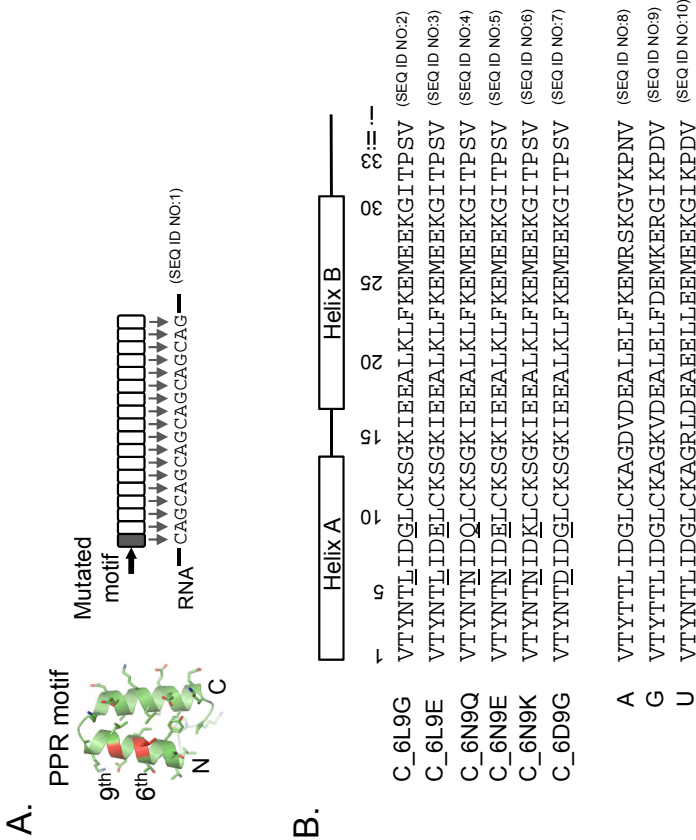
[Claim 17]

A method for manipulating a nucleic acid, which uses the polypeptide according to any one of claims 1 to 3, the protein according to any one of claims 6 to 10, or the vector according to claim 15 (implementation in human individual is excluded).

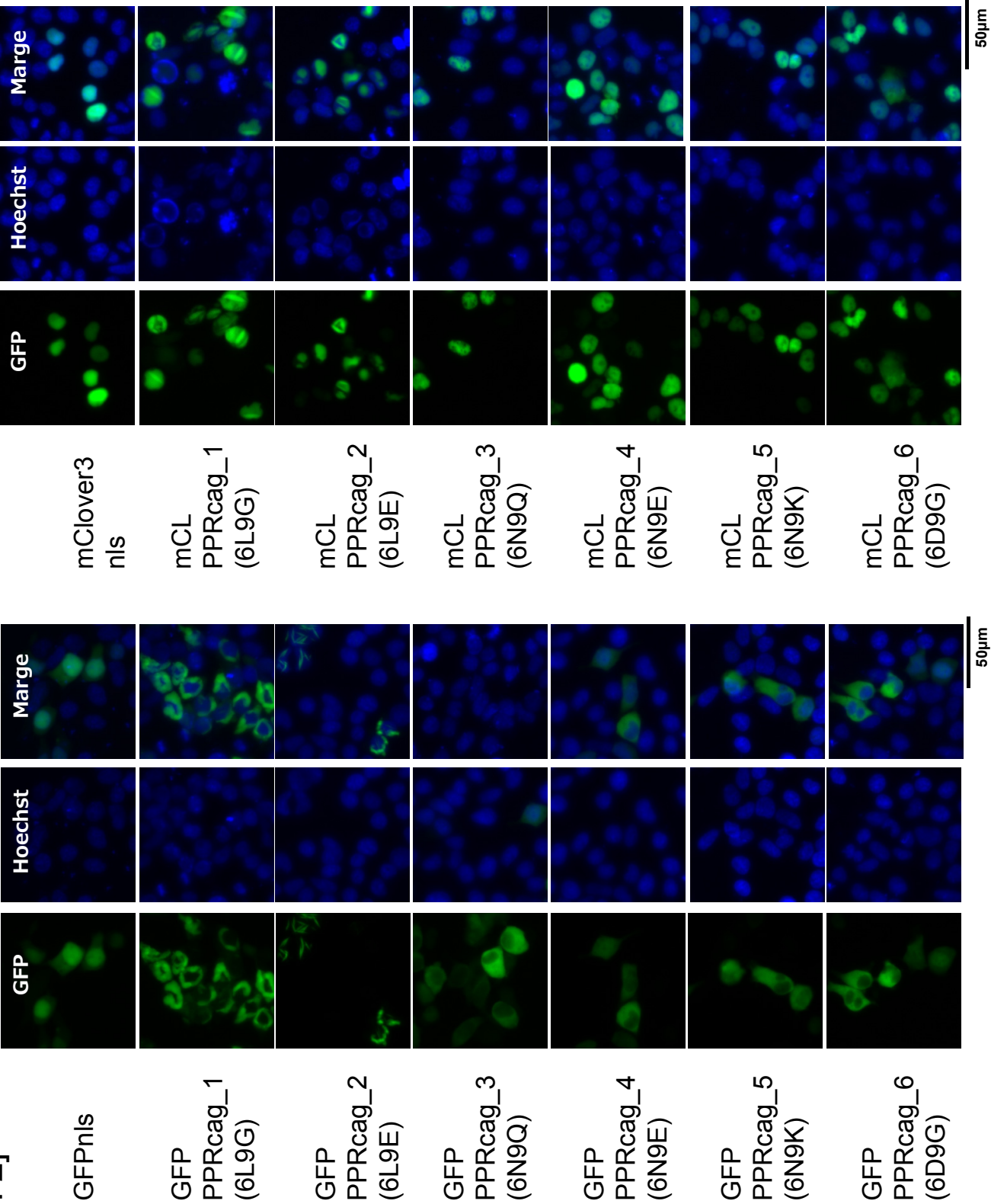
[Claim 18]

A method for producing an organism, which comprises the manipulation method according to claim 17.

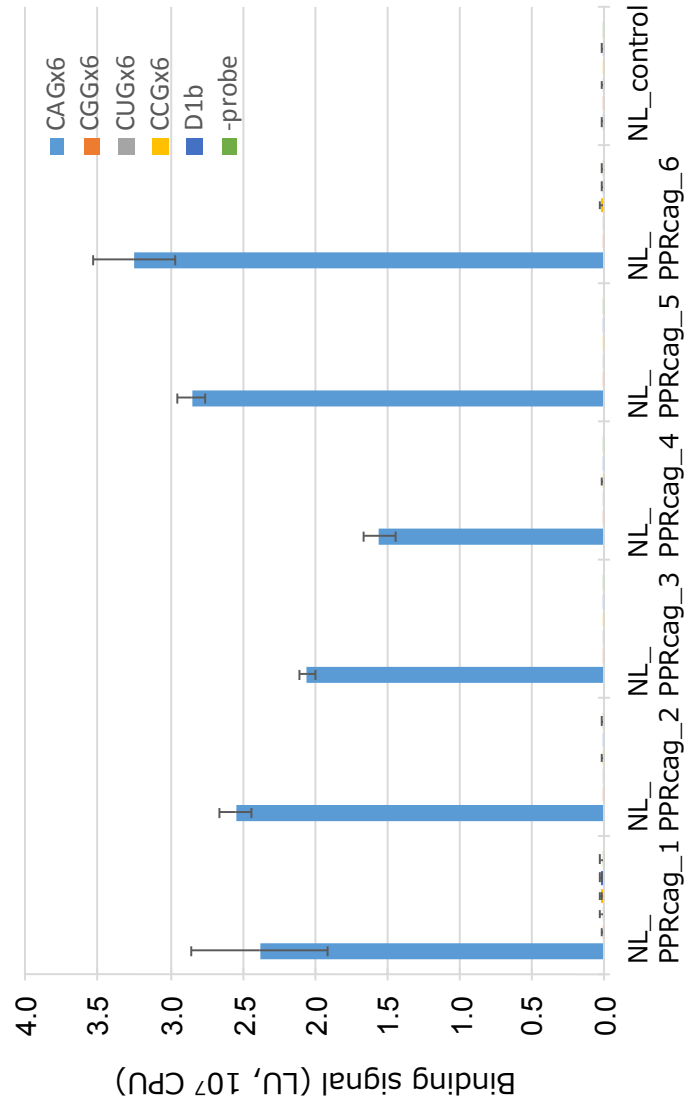
[Fig. 1]



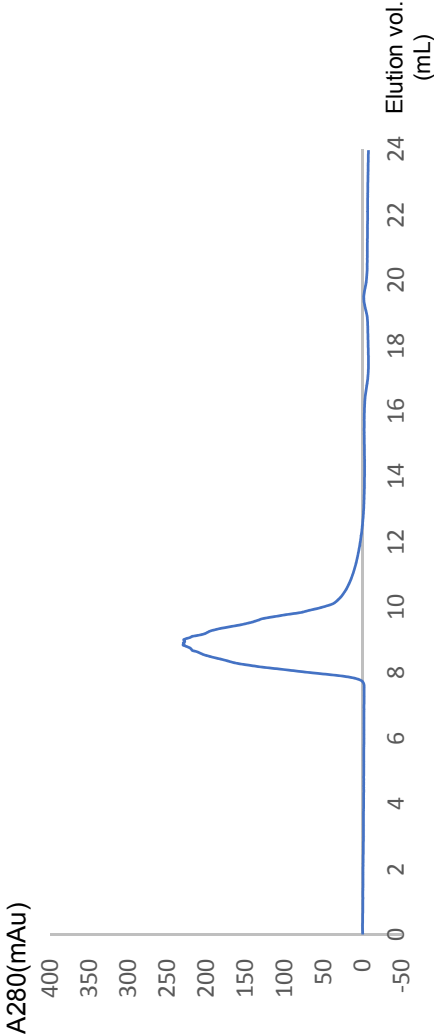
[Fig. 2]



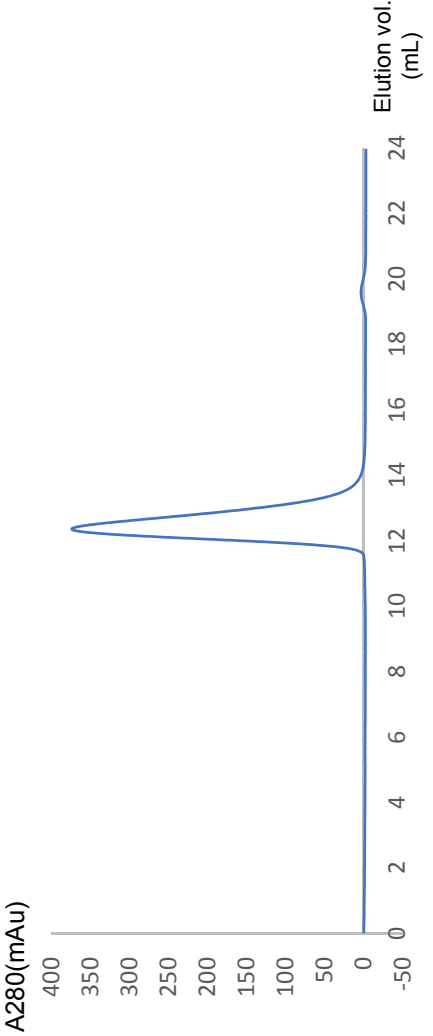
[Fig. 3]



[Fig. 4]



v2



v3.2

1  
SEQUENCE LISTING

<110> EditForce Inc.  
Kyushu University

<120> Low-cohesive PPR Proteins and Use thereof

<130> F20689K

<150> JP2019-100553

<151> 2019-05-29

<160> 64

<170> PatentIn version 3.5

<210> 1

<211> 18

<212> DNA

<213> Artificial sequence

<220>

<223> target sequence

<400> 1

cagcagcagc agcagcag

18

<210> 2

<211> 35

<212> PRT

<213> Artificial sequence

<220>

<223> PPR motif

<400> 2

Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile  
1 5 10 15

Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr  
20 25 30

Pro Ser Val  
35

<210> 3

<211> 35

<212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 3

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Thr | Tyr | Asn | Thr | Leu | Ile | Asp | Glu | Leu | Cys | Lys | Ser | Gly | Lys | Ile |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Glu | Ala | Leu | Lys | Leu | Phe | Lys | Glu | Met | Glu | Glu | Lys | Gly | Ile | Thr |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

|     |     |     |
|-----|-----|-----|
| Pro | Ser | Val |
|     |     | 35  |

<210> 4  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif, 1st\_U (v3.2\_U)

<400> 4

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Thr | Tyr | Asn | Thr | Asn | Ile | Asp | Gln | Leu | Cys | Lys | Ser | Gly | Lys | Ile |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Glu | Ala | Leu | Lys | Leu | Phe | Lys | Glu | Met | Glu | Glu | Lys | Gly | Ile | Thr |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

|     |     |     |
|-----|-----|-----|
| Pro | Ser | Val |
|     |     | 35  |

<210> 5  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 5

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Thr | Tyr | Asn | Thr | Asn | Ile | Asp | Glu | Leu | Cys | Lys | Ser | Gly | Lys | Ile |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

|   |   |         |    |
|---|---|---------|----|
| 1 | 5 | 3<br>10 | 15 |
|---|---|---------|----|

Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr  
                   20                  25                  30

Pro Ser Val  
           35

<210> 6  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 6

|                                                                 |
|-----------------------------------------------------------------|
| Val Thr Tyr Asn Thr Asn Ile Asp Lys Leu Cys Lys Ser Gly Lys Ile |
| 1                  5                  10                  15    |

Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr  
                   20                  25                  30

Pro Ser Val  
           35

<210> 7  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 7

|                                                                 |
|-----------------------------------------------------------------|
| Val Thr Tyr Asn Thr Asp Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile |
| 1                  5                  10                  15    |

Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr  
                   20                  25                  30

Pro Ser Val  
           35

<210> 8  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 8

Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Asp Val  
 1 5 10 15

Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys  
 20 25 30

Pro Asn Val  
 35

<210> 9  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 9

Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val  
 1 5 10 15

Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys  
 20 25 30

Pro Asp Val  
 35

<210> 10  
 <211> 35  
 <212> PRT  
 <213> Artificial sequence

<220>  
 <223> PPR motif

<400> 10

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Thr | Tyr | Asn | Thr | Leu | Ile | Asp | Gly | Leu | Cys | Lys | Ala | Gly | Arg | Leu |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Glu | Ala | Glu | Glu | Leu | Leu | Glu | Glu | Met | Glu | Glu | Lys | Gly | Ile | Lys |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |

|     |     |     |
|-----|-----|-----|
| Pro | Asp | Val |
|     |     | 35  |

<210> 11

<211> 2303

<212> DNA

<213> Artificial sequence

<220>

<223> PPR gene

<220>

<221> misc\_feature

<222> (1894)..(1894)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (1935)..(1935)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (1963)..(1963)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2004)..(2004)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2032)..(2032)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2073)..(2073)

<223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2101)..(2101)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2142)..(2142)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2170)..(2170)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2211)..(2211)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2239)..(2239)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2280)..(2280)  
 <223> n is a, c, g, t or u

<400> 11  
 gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg 60  
 aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca 120  
 ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag 180  
 atgcgaggca agggcgtgaa gccaacgtg gtgacataca ccaccctgat cgacggcctg 240  
 tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc 300  
 atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag 360  
 atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg 420  
 gttacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 480  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc 540  
 ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag 600  
 atgaaggaga ggggcatcaa gcccgcgtg gttacataca acaccctgat cgacggcctg 660

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 720  |
| atcaccccca gcgtggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac  | 780  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg   | 840  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 900  |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtggtcac atacaacacc  | 960  |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 1020 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg   | 1080 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1140 |
| gtgaagccca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag  | 1200 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgacgtg  | 1260 |
| gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1320 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggtcac atacaccaca  | 1380 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag  | 1440 |
| atgcggagca agggcgtgaa gcccacgtg gttacataca ccaccctgat cgacggcctg   | 1500 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc  | 1560 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag  | 1620 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg   | 1680 |
| gtgacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 1740 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc  | 1800 |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag  | 1860 |
| atgaaggaga ggggcatcaa gcccgacgag vtyntdgcks gkakkmgts vtytttdgck   | 1920 |
| agdvdakmrs kgvknvtyt tdgckagkvd admkrkdvv tyntdgcksg kakkmgtsv     | 1980 |
| vtytttdgcka gdvdakmrsk gvknvtytt dgckagkvda dmkrkdvvt yntdgcksgk   | 2040 |
| akkmkgtsvv tytttdgckag dvdakmrskg vknvtytttd gckagkvdad mkrkdvvtty | 2100 |
| ntdgcksgka kkmkgtsvvt ytttdgckagd vdakmrskgv knvtytttdg ckagkvdadm | 2160 |
| krkdvvtyn tdgcksgkak kkmkgtsvvy ttdgckagdv dakmrskgvk nvtytttdgc   | 2220 |

kagkvdadmk rgkdvvtynt dgcksgkakk mkgtsvvtyt tdgckagdvd akmrskgvkn 2280

vvtyttgdck agkvdadmkr gkd 2303

<210> 12  
 <211> 2302  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR gene

<220>  
 <221> misc\_feature  
 <222> (1894)..(1894)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1934)..(1934)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1962)..(1962)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2003)..(2003)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2031)..(2031)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2072)..(2072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2100)..(2100)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2141)..(2141)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2169)..(2169)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2210)..(2210)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2238)..(2238)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2279)..(2279)  
 <223> n is a, c, g, t or u

<400> 12  
 gtcacatata acaccctgat cgacgaactg tgcaagagcg gcaagatcga ggaggccctg 60  
 aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca 120  
 ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag 180  
 atgcggagca agggcgtgaa gcccacgtg gtgacatata ccaccctgat cgacggcctg 240  
 tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc 300  
 atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag 360  
 atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg 420  
 gttacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 480  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc 540  
 ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag 600  
 atgaaggaga ggggcatcaa gcccgcgtg gttacatata acaccctgat cgacggcctg 660  
 tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc 720  
 atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac 780  
 gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg 840  
 gtgacatata ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg 900  
 gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtgggtcac atacaacacc 960

|                                                                   |      |
|-------------------------------------------------------------------|------|
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag | 1020 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg  | 1080 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc | 1140 |
| gtgaagccca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag | 1200 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg  | 1260 |
| gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg | 1320 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggtcac atacaccaca | 1380 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag | 1440 |
| atgcggagca agggcgtgaa gccaacgtg gttacataca ccaccctgat cgacggcctg  | 1500 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc | 1560 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag | 1620 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg  | 1680 |
| gtgacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc | 1740 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc | 1800 |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag | 1860 |
| atgaaggaga ggggcatcaa gcccgcagag vtyntdcksg kakkmkgtsv vtyttgdcka | 1920 |
| gdvdakmrsk gvknvtytt dgckagkvda dmkrkgdvvt yntdgcksgk akkmkgtsvv  | 1980 |
| tyttgdckag dvdakmrskg vknvtyttgd gckagkvdad mkrkgdvvt yntdgcksgka | 2040 |
| kkmkgtsvvt yttgdckagd vdakmrskgv knvtyttgd ckagkvdadm krgkdvvtyn  | 2100 |
| tdgcksgkak kmkgtsvvty ttdgckagdv dakmrskgvk nvtyttgdgc kagkvdadmk | 2160 |
| rgkdvvtynt dgcksgkak mkgtsvvtyt tdgckagdvd akmrskgvkn vtyttgdgck  | 2220 |
| agkvdadmkrg kvdvvtyntd gcksgkakkm kgtsvvtytt dgckagdvd kmrskgvknv | 2280 |
| vtyttgdcka gkvdadmkrg kd                                          | 2302 |

<210> 13  
 <211> 2303  
 <212> DNA  
 <213> Artificial sequence

<220>  
<223> PPR gene

<220>  
<221> misc\_feature  
<222> (1894)..(1894)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (1896)..(1896)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (1935)..(1935)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (1963)..(1963)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2004)..(2004)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2032)..(2032)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2073)..(2073)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2101)..(2101)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2142)..(2142)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2170)..(2170)  
<223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2211)..(2211)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2239)..(2239)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2280)..(2280)  
 <223> n is a, c, g, t or u

<400> 13  
 gtcacatata acaccaacat cgaccagctg tgcaagagcg gcaagatcga ggaggccctg 60  
 aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca 120  
 ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag 180  
 atgctggagca agggcggtgaa gcccacgtg gtgacatata ccaccctgat cgacggcctg 240  
 tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc 300  
 atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag 360  
 atcgaggagg ccctgaagct gttcaaaggag atggaggaga agggcatcac cccagcgtg 420  
 gttacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 480  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc 540  
 ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag 600  
 atgaaggaga ggggcatcaa gcccgcgtg gttacatata acaccctgat cgacggcctg 660  
 tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc 720  
 atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac 780  
 gtggacgaag ccctcgagct gttcaaagag atgctggagca agggcggtgaa gcccacgtg 840  
 gtgacatata ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg 900  
 gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtgggtcac atacaacacc 960  
 ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaaggag 1020  
 atggaggaga agggcatcac cccagcgtg gttacatata ccacactgat cgacggactg 1080

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1140 |
| gtgaagccca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag  | 1200 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgctg  | 1260 |
| gtcacatata acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1320 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggtcac atacaccaca  | 1380 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag  | 1440 |
| atgcggagca agggcgtgaa gcccacgtg gttacatata ccaccctgat cgacggcctg   | 1500 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc  | 1560 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag  | 1620 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac ccccgagctg  | 1680 |
| gtgacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 1740 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc  | 1800 |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag  | 1860 |
| atgaaggaga ggggcatcaa gcccgcgag vtyntndcks gkakkmgts vtyttgdgck    | 1920 |
| agdvdakmrs kgvknvtyt tdgckagkvd admkrkgdv tyntdgcksg kakkmgtsv     | 1980 |
| vtyttgdgcka gdvdakmrsk gvknvtytt dgckagkvda dmkrkgdvvt yntdgcksgk  | 2040 |
| akkmkgtsvv tyttgdgckag dvdakmrskg vknvtyttg gckagkvdad mkrkgdvty   | 2100 |
| ntdgcksgka kkmkgtsvvt yttgdgckagd vdakmrskgv knvtyttgdg ckagkvdadm | 2160 |
| krkgdvvtyn tdgcksgkak kmkgtsvty ttdgckagdv dakmrskgvk nvtyttgdg    | 2220 |
| kagkvdadmk rgkdvvtynt dgcksgkakk mkgtsvtyt tdgckagdvd akmrskgvkn   | 2280 |
| vtyttgdgck agkvdadmkr gkd                                          | 2303 |

<210> 14  
 <211> 2303  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR gene

<220>

<221> misc\_feature  
 <222> (1894)..(1894)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1896)..(1896)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1935)..(1935)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1963)..(1963)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2004)..(2004)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2032)..(2032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2073)..(2073)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2101)..(2101)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2142)..(2142)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2170)..(2170)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2211)..(2211)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2239)..(2239)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2280)..(2280)  
 <223> n is a, c, g, t or u

<400> 14  
 gtcacataca acaccaacat cgacgaactg tgcaagagcg gcaagatcga ggaggccctg 60  
 aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggtcac atacaccaca 120  
 ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag 180  
 atgctggagca agggcgtgaa gcccacgtg gtgacataca ccaccctgat cgacggcctg 240  
 tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc 300  
 atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag 360  
 atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg 420  
 gttacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 480  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc 540  
 ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag 600  
 atgaaggaga ggggcatcaa gcccgcgtg gttacataca acaccctgat cgacggcctg 660  
 tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc 720  
 atcaccccca gcgtggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac 780  
 gtggacgaag ccctcgagct gttcaaagag atgctggagca agggcgtgaa gcccacgtg 840  
 gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg 900  
 gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtggtcac atacaacacc 960  
 ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag 1020  
 atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg 1080  
 tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc 1140  
 gtgaagccca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag 1200  
 gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg 1260

```

gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg      1320
aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca      1380
ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag      1440
atgcggagca agggcgtgaa gccaacgtg gttacataca ccaccctgat cgacggcctg      1500
tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagagggggc      1560
atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag      1620
atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg      1680
gtgacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc      1740
gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc      1800
ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag      1860
atgaaggaga ggggcatcaa gcccgcagag vtyntndcks gkakkmgts vvttytdgck      1920
agdvda kmrs kgvknvtyt tdgckagkvd admkr gkdvv tyntdgcksg kakkmgtsv      1980
vtytdgcka gdvdakmrsk gvknvtytt dgckagkvda dmkr gkdvvt yntdgcksgk      2040
akkmkgtsvv tytdgckag dvdakmrskg vknvtytt dgckagkvdad mkr gkdvvt y      2100
ntdgcksgka kkmkgtsvvt ytdgckagd vdakmrskgv knvtytt dg ckagkvdad m      2160
kr gkdvvtyn tdgcksgkak kmkgtsvvy ttdgckagdv dakmrskgvk nvtytt dg c      2220
kagkvdadmk rgkdvvtynt dgcksgkakk mkgtsvvyt tdgckagdvd akmrskgvkn      2280
vvttytdgck agkvdadmk r gkd      2303

```

```

<210> 15
<211> 2304
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> PPR gene

```

```

<220>
<221> misc_feature
<222> (1894)..(1894)
<223> n is a, c, g, t or u

```

```

<220>

```

<221> misc\_feature  
 <222> (1896)..(1896)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1936)..(1936)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1964)..(1964)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2005)..(2005)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2033)..(2033)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2074)..(2074)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2102)..(2102)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2143)..(2143)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2171)..(2171)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2212)..(2212)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2240)..(2240)  
 <223> n is a, c, g, t or u

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (2281)..(2281)

&lt;223&gt; n is a, c, g, t or u

&lt;400&gt; 15

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gtcacatata acaccaacat cgacaaactg tgcaagagcg gcaagatcga ggaggccctg  | 60   |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca | 120  |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag  | 180  |
| atgcggagca agggcgtgaa gcccacgtg gtgacatata ccaccctgat cgacggcctg   | 240  |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc  | 300  |
| atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag | 360  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg   | 420  |
| gttacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 480  |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc | 540  |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag  | 600  |
| atgaaggaga ggggcatcaa gcccgcgtg gttacatata acaccctgat cgacggcctg   | 660  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 720  |
| atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac | 780  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg   | 840  |
| gtgacatata ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 900  |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtgggtcac atacaacacc | 960  |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 1020 |
| atggaggaga agggcatcac cccagcgtg gttacatata ccacactgat cgacggactg   | 1080 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1140 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag | 1200 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg   | 1260 |
| gtcacatata acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1320 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca | 1380 |

```

ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag      1440
atgcgaggagca agggcggtgaa gccaacgtg gttacataca ccaccctgat cgacggcctg      1500
tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc      1560
atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag      1620
atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg      1680
gtgacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc      1740
gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc      1800
ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag      1860
atgaaggaga ggggcatcaa gcccgcgag vtyntndkck sgkakkmkg tsvtytttdgc      1920
kagdvdamr skgvknvty ttdgckagkv dadmkrkdv vtyntdgcks gkakkmkgts      1980
vtytttdgck agdvdamrs kgvknvtyt tdgckagkv admkrkdvv tyntdgcksg      2040
kakkmkgtsv vtytttdgcka gdvdakmrsk gvknvtytt dgckagkvda dmkrkdvvt      2100
yntdgcksgk akkmkgtsvv tytttdgckag dvdakmrskg vknvtytttd gckagkvdad      2160
mkrkdvvty ntdgcksgka kkmkgtsvvt ytttdgckagd vdakmrskgv knvtytttdg      2220
ckagkvdadm krgkdvvty ntdgcksgkak kmkgtsvty ttdgckagdv dakmrskgvk      2280
nvtytttdgc kagkvdadmk rgkd      2304

```

```

<210> 16
<211> 2304
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> PPR gene

```

```

<220>
<221> misc_feature
<222> (1894)..(1894)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (1936)..(1936)
<223> n is a, c, g, t or u

```

```

<220>

```

<221> misc\_feature  
 <222> (1964)..(1964)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2005)..(2005)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2033)..(2033)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2074)..(2074)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2102)..(2102)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2143)..(2143)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2171)..(2171)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2212)..(2212)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2240)..(2240)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2281)..(2281)  
 <223> n is a, c, g, t or u

<400> 16  
 gtcacatata acaccgatat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg 60  
 aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca 120

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag   | 180  |
| atgcgaggagca agggcggtgaa gccaacgtg gtgacataca ccaccctgat cgacggcctg | 240  |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc   | 300  |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag   | 360  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg    | 420  |
| gttacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc   | 480  |
| gagctgttca aagagatgcg gagcaagggc gtgaagcca acgtggtcac atacaccacc    | 540  |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag   | 600  |
| atgaaggaga ggggcatcaa gcccgcgtg gttacataca acaccctgat cgacggcctg    | 660  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc   | 720  |
| atcaccccca gcgtggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac   | 780  |
| gtggacgaag ccctcgagct gttcaaagag atgcgaggca agggcggtgaa gccaacgtg   | 840  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg   | 900  |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtggtcac atacaacacc   | 960  |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag   | 1020 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg    | 1080 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc   | 1140 |
| gtgaagcca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag    | 1200 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg    | 1260 |
| gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg   | 1320 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggtcac atacaccaca   | 1380 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag   | 1440 |
| atgcgaggagca agggcggtgaa gccaacgtg gttacataca ccaccctgat cgacggcctg | 1500 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc   | 1560 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag   | 1620 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg    | 1680 |

```

gtgacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 1740
gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtcac atacaccacc 1800
ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag 1860
atgaaggaga ggggcatcaa gcccgcacgag vtyntddgck sgkakkmkgt svvtyttddgc 1920
kagdvda kmr skgvknv vty ttdgckagkv dadmkr gkdv vtyntdgcks gkakkmkgts 1980
vvtyttddgck agdvda kmrs kgvknv vtyt tdgckagkv d admkr gkdvv tyntdgcksg 2040
kakkmkgtsv vtyttddgcka gdvdakmrsk gvk n vtytt dgckagkvda dmkr gkdvv t 2100
yntdgcksgk akkmkgtsvv tyttddgckag dvdakmrskg vknv vtytt d gckagkvdad 2160
mkr gkdvvty ntdgcksgka kkmkgtsvvt yttddgckagd vdakmrskgv knv vtytt dg 2220
ckagkvdad m krgkdvvty n tdgcksgkak kmkgtsv vty ttdgckagdv dakmrskgvk 2280
nvvtyttddgc kagkvdadmk rgkd 2304

```

```

<210> 17
<211> 3481
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> PPR vector

```

```

<220>
<221> misc_feature
<222> (2858)..(2858)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2914)..(2914)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2926)..(2926)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2934)..(2934)
<223> n is a, c, g, t or u

```

```

<220>

```

<221> misc\_feature  
 <222> (2939)..(2939)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2941)..(2941)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2944)..(2944)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2956)..(2956)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2960)..(2960)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2969)..(2969)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2976)..(2976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2986)..(2986)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3070)..(3070)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3098)..(3098)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3139)..(3139)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3167)..(3167)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3208)..(3208)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3236)..(3236)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3277)..(3277)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3305)..(3305)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3346)..(3346)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3374)..(3374)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (3415)..(3415)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3441)..(3441)

<223> n is a, c, g, t or u

<400> 17

|                                                                   |      |
|-------------------------------------------------------------------|------|
| atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggctgag | 60   |
| ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc | 120  |
| acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg | 180  |
| cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc | 300  |
| atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac | 360  |
| accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg | 420  |
| gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag | 480  |
| aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag | 540  |
| ctcgccgacc actaccagca gaacaccccc atcggcgacg gccccgtgtc gctgcccgac | 600  |
| aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac | 660  |
| atggtcctgc tggagttcgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac | 720  |
| aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat  | 780  |
| atgggaaact ccgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag | 840  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg  | 900  |
| gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc | 960  |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtgac atacaccacc | 1020 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag | 1080 |
| atgaaggaga ggggcatcaa gcccgcgtg gtcacataca acaccctgat cgacggcctg  | 1140 |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc | 1200 |
| atcccccca gcgtggttac atacaccaca ctgatcgacg gactgtgtaa agccggcgac  | 1260 |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccaacgtg | 1320 |

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| gtcacataca | ccaccctgat | cgacggcctg | tgcaaggccg | gcaaggtgga | tgaggccctg | 1380 |
| gagctgttcg | acgagatgaa | ggagaggggc | atcaagcccg | acgtggttac | atacaacacc | 1440 |
| ctgatcgacg | gcctgtgcaa | gagcggcaag | atcgaggagg | ccctgaagct | gttcaaggag | 1500 |
| atggaggaga | agggcacac  | ccccagcgtg | gtcacataca | ccacactgat | cgacggactg | 1560 |
| tgtaaagccg | gcgacgtgga | cgaagccctc | gagctgttca | aagagatgcg | gagcaagggc | 1620 |
| gtgaagccca | acgtggtgac | atacaccacc | ctgatcgacg | gcctgtgcaa | ggccggcaaa | 1680 |
| gtggacgagg | ccctggagct | gttcgacgag | atgaaggaga | ggggcatcaa | gcccgcgctg | 1740 |
| gtcacataca | acaccctgat | cgacggcctg | tgcaagagcg | gcaagatcga | ggaggccctg | 1800 |
| aagctgttca | aggagatgga | ggagaagggc | atcaccccca | gcgtggttac | atacaccaca | 1860 |
| ctgatcgacg | gactgtgtaa | agccggcgac | gtggacgaag | ccctcgagct | gttcaaagag | 1920 |
| atgcggagca | agggcgtgaa | gccaacgtg  | gtcacataca | ccaccctgat | cgacggcctg | 1980 |
| tgcaaggccg | gcaaggtgga | tgaggccctg | gagctgttcg | acgagatgaa | ggagaggggc | 2040 |
| atcaagcccg | acgtggtcac | atacaacacc | ctgatcgacg | gcctgtgcaa | gagcggcaag | 2100 |
| atcgaggagg | ccctgaagct | gttcaaggag | atggaggaga | agggcacac  | ccccagcgtg | 2160 |
| gtcacataca | ccacactgat | cgacggactg | tgtaaagccg | gcgacgtgga | cgaagccctc | 2220 |
| gagctgttca | aagagatgcg | gagcaagggc | gtgaagccca | acgtggttac | atacaccacc | 2280 |
| ctgatcgacg | gcctgtgcaa | ggccggcaaa | gtggacgagg | ccctggagct | gttcgacgag | 2340 |
| atgaaggaga | ggggcatcaa | gcccgcgctg | gtcacataca | acaccctgat | cgacggcctg | 2400 |
| tgcaagagcg | gcaagatcga | ggaggccctg | aagctgttca | aggagatgga | ggagaagggc | 2460 |
| atcaccccca | gcgtggtgac | atacaccaca | ctgatcgacg | gactgtgtaa | agccggcgac | 2520 |
| gtggacgaag | ccctcgagct | gttcaaagag | atgcggagca | agggcgtgaa | gccaacgtg  | 2580 |
| gtcacataca | ccaccctgat | cgacggcctg | tgcaaggccg | gcaaggtgga | tgaggccctg | 2640 |
| gagctgttcg | acgagatgaa | ggagaggggc | atcaagcccg | acgagctgac | ctacaacacc | 2700 |
| ctgatcagcg | gcctgggcaa | ggccggcaga | gccagagacc | ccccgtgct  | cagtagcggg | 2760 |
| gactataagg | accacgacgg | agactacaag | gatcatgata | ttgattacaa | agacgatgac | 2820 |
| gataagatgg | ccggacgcta | gmagvskgtg | vvvdgdvngh | ksvsgggdat | ygktkcttgk | 2880 |
| vwtvtttygv | csrydhmkhd | ksamgyvrtk | ddgnyktrav | kgdtvnrkgd | kdgngkhyny | 2940 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm   | 3000 |
| dykkkkkrkvs hgsggsgggh mgnsvvtynt dgcksgkakk mkgtsvvtyt tdgckagdvd | 3060 |
| akmrskgvkn vvytttdgck agkvdadmk r gkdvvtyntd gcksgkakkm kgtsvvtytt | 3120 |
| dgckagdvd kmrskgvknv vtytttdgcka gkvdadmkrg kdvvtynldg cksgkakkmk  | 3180 |
| gtsvvtytttd gckagdvdak mrskgvknv tytttdgckag kvdadmkrgk dvvtynldgc | 3240 |
| ksgkakkmkg tsvvytttdg ckagdvdakm rskgvknvvt ytttdgckagk vdadmkrgkd | 3300 |
| vvtyntdgck sgkakkmkg svvytttdgc kagdvdakmr skgvknvty ttdgckagkv    | 3360 |
| dadmkrgkdv vtyntdgcks gkakkmkgts vvytttdgck agdvdakmrs kgvknvtyt   | 3420 |
| tdgckagkvd admkrgkdy ntsggkagra rdvssgdykd hdgdykdhd ykdddkmag     | 3480 |
| r                                                                  | 3481 |

<210> 18  
 <211> 3480  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2858)..(2858)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2914)..(2914)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2934)..(2934)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (2939)..(2939)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2941)..(2941)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2944)..(2944)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2956)..(2956)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2960)..(2960)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2969)..(2969)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2976)..(2976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2986)..(2986)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3069)..(3069)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3097)..(3097)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3138)..(3138)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3166)..(3166)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3207)..(3207)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3235)..(3235)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3276)..(3276)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3304)..(3304)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3345)..(3345)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3373)..(3373)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3414)..(3414)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3440)..(3440)

<223> n is a, c, g, t or u

<400> 18

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggctgag  | 60   |
| ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc  | 120  |
| acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg  | 180  |
| cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac  | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc  | 300  |
| atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac  | 360  |
| accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg  | 420  |
| gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag  | 480  |
| aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag  | 540  |
| ctcgccgacc actaccagca gaacaccccc atcggcgacg gccccgtgtc gctgcccgac  | 600  |
| aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac  | 660  |
| atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac  | 720  |
| aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat   | 780  |
| atgggaaact ccgtgggtcac atacaacacc ctgatcgacg aactgtgcaa gagcggcaag | 840  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg   | 900  |
| gtcacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 960  |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtgac atacaccacc | 1020 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag  | 1080 |
| atgaaggaga ggggcatcaa gcccgcgtg gtcacatata acaccctgat cgacggcctg   | 1140 |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 1200 |
| atcaccccca gcgtgggttac atacaccaca ctgatcgacg gactgtgtaa agccggcgac | 1260 |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccaacgtg  | 1320 |
| gtcacatata ccaccctgat cgacggcctg tgcaaggccg gcaaggtgga tgaggccctg  | 1380 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| gagctgttcg acgagatgaa ggagagggggc atcaagcccg acgtgggttac atacaacacc | 1440 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag   | 1500 |
| atggaggaga agggcatcac cccagcgtg gtcacataca ccacactgat cgacggactg    | 1560 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc   | 1620 |
| gtgaagccca acgtggtgac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaaa   | 1680 |
| gtggacgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg    | 1740 |
| gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg   | 1800 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggttac atacaccaca  | 1860 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag   | 1920 |
| atgcggagca agggcgtgaa gccaacgtg gtcacataca ccaccctgat cgacggcctg    | 1980 |
| tgcaaggccg gcaaggtgga tgaggccctg gagctgttcg acgagatgaa ggagaggggc   | 2040 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag   | 2100 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg    | 2160 |
| gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc   | 2220 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggttac atacaccacc  | 2280 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag   | 2340 |
| atgaaggaga ggggcatcaa gcccgcgtg gtcacataca acaccctgat cgacggcctg    | 2400 |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc   | 2460 |
| atcaccccca gcgtggtgac atacaccaca ctgatcgacg gactgtgtaa agccggcgac   | 2520 |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gccaacgtg    | 2580 |
| gtcacataca ccaccctgat cgacggcctg tgcaaggccg gcaaggtgga tgaggccctg   | 2640 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgagctgac ctacaacacc   | 2700 |
| ctgatcagcg gcctgggcaa ggccggcaga gccagagacc ccccgctgct cagtagcggg   | 2760 |
| gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac   | 2820 |
| gataagatgg ccggacgcta gmagvsktg vvdgdvngh ksvsggdat ygktkcttgk      | 2880 |
| vwtvtttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtnrkgd kdgnghkyny    | 2940 |

nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm 3000  
 dykkkkkrkvs hgsggsgggh mgnsvvtynt dcksgkakkm kgtsvvtytt dgckagdvd 3060  
 kmrskgvknv vtyttgdcka gkvdadmkrk kdvvtyntdg cksgkakkmk gtsvvtytt 3120  
 gckagdvdak mrskgvknv tyttgdckag kvdadmkrk dvvtyntdgc ksgkakkmkg 3180  
 tsvvtyttgd ckagdvdakm rskgvknv yttdgckag vdadmkrkd vvtyntdgck 3240  
 sgkakkmkg svvtyttgd kagdvdakmr skgvknv tyttgdckag kv dadmkrkd 3300  
 vtyntdgcks gkakkmkgts vvtyttgdck agdvadkrmr kgvknvtyt tdgckagkv 3360  
 admkrkdvv tyntdgcksg kakkmkgts vtyttgdcka gvdakmrsk gvknvtytt 3420  
 dgckagkvda dmkrkdtynt tsggkagnar dvssgdykdh dgdykdhdy kdddkmagr 3480

<210> 19  
 <211> 3481  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2858)..(2858)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2914)..(2914)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2934)..(2934)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2939)..(2939)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2941)..(2941)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2944)..(2944)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2956)..(2956)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2960)..(2960)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2969)..(2969)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2976)..(2976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2986)..(2986)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3031)..(3031)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3070)..(3070)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3098)..(3098)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3139)..(3139)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3167)..(3167)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3208)..(3208)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3236)..(3236)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3277)..(3277)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3305)..(3305)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3346)..(3346)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3374)..(3374)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (3415)..(3415)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3441)..(3441)  
 <223> n is a, c, g, t or u

<400> 19  
 atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccat cctggtcgag 60  
 ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc 120  
 acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg 180  
 cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac 240  
 atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc 300  
 atctttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac 360  
 accctgggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg 420  
 gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag 480  
 aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag 540  
 ctgcccgaac actaccagca gaacaccccc atcggcgacg gccccgtgtc gctgcccgaac 600  
 aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac 660  
 atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac 720  
 aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat 780  
 atgggaaact ccgtgggtcac atacaacacc aacatcgacc agctgtgcaa gagcggcaag 840  
 atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg 900  
 gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 960  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtgac atacaccacc 1020  
 ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag 1080  
 atgaaggaga ggggcatcaa gcccgcgtg gtcacataca acaccctgat cgacggcctg 1140  
 tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc 1200  
 atcccccca gcgtgggttac atacaccaca ctgatcgacg gactgtgtaa agccggcgac 1260  
 gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg 1320

|                                                                   |      |
|-------------------------------------------------------------------|------|
| gtcacataca ccaccctgat cgacggcctg tgcaaggccg gcaaggtgga tgaggccctg | 1380 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtggttac atacaacacc | 1440 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag | 1500 |
| atggaggaga agggcatcac cccagcgtg gtcacataca ccacactgat cgacggactg  | 1560 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc | 1620 |
| gtgaagccca acgtggtgac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaaa | 1680 |
| gtggacgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg  | 1740 |
| gtcacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg | 1800 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtggttac atacaccaca | 1860 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag | 1920 |
| atgcggagca agggcgtgaa gccaacgtg gtcacataca ccaccctgat cgacggcctg  | 1980 |
| tgcaaggccg gcaaggtgga tgaggccctg gagctgttcg acgagatgaa ggagaggggc | 2040 |
| atcaagcccg acgtggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag | 2100 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg  | 2160 |
| gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc | 2220 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggttac atacaccacc | 2280 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag | 2340 |
| atgaaggaga ggggcatcaa gcccgcgtg gtcacataca acaccctgat cgacggcctg  | 2400 |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc | 2460 |
| atcaccccca gcgtggtgac atacaccaca ctgatcgacg gactgtgtaa agccggcgac | 2520 |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gccaacgtg  | 2580 |
| gtcacataca ccaccctgat cgacggcctg tgcaaggccg gcaaggtgga tgaggccctg | 2640 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgagctgac ctacaacacc | 2700 |
| ctgatcagcg gcctgggcaa ggccggcaga gccagagacc ccccgtgct cagtagcggg  | 2760 |
| gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac | 2820 |
| gataagatgg ccggacgcta gmagvsktg vvdgdvngh ksvsggdat ygktkcttgk    | 2880 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| vwtvtttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtvnrkgd kdgnghkyny   | 2940 |
| nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm    | 3000 |
| dykkkkkrkvs hgsggsggggh mgnsvvtynt ndcksgkakk mkgtsvvtyt tdgckagdvd | 3060 |
| akmrskgvkn vvttyttgck agkvdadmkr gkdvvtyntd gcksgkakkm kgtsvvtytt   | 3120 |
| dgckagdvd kmrskgvknv vtyttgcka gkvdadmkr gkdvvtyntdg cksgkakkmk     | 3180 |
| gtsvvtyttg gckagdvdak mrskgvknv tyttgckag kvdadmkr gkdvvtyntdg      | 3240 |
| ksgkakkmkg tsvvtyttg ckagdvdakm rskgvknv yttdgckag vdadmkr gkd      | 3300 |
| vvtyntdgck sgkakkmkg svvtyttg kagdvdakmr skgvknv tyttgckagkv        | 3360 |
| dadmkr gkdv tyntdgcks gkakkmkgts vvttyttgck agdvdakmr kgvknv tyt    | 3420 |
| tdgckagkv admkr gkdty ntsggkagra rdvssgdykd hdgdykd hdd ykddddk mag | 3480 |
| r                                                                   | 3481 |

<210> 20  
 <211> 3481  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2858)..(2858)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2914)..(2914)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2934)..(2934)  
 <223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (2939)..(2939)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2941)..(2941)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2944)..(2944)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2956)..(2956)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2960)..(2960)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2969)..(2969)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2976)..(2976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2986)..(2986)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3031)..(3031)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3070)..(3070)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3098)..(3098)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3139)..(3139)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3167)..(3167)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3208)..(3208)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3236)..(3236)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3277)..(3277)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3305)..(3305)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3346)..(3346)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (3374)..(3374)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3415)..(3415)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3441)..(3441)

<223> n is a, c, g, t or u

<400> 20

|                                                                   |      |
|-------------------------------------------------------------------|------|
| atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccac cctggtcgag | 60   |
| ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc | 120  |
| acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg | 180  |
| cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc | 300  |
| atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac | 360  |
| accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg | 420  |
| gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag | 480  |
| aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag | 540  |
| ctcgccgacc actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgac | 600  |
| aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac | 660  |
| atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac | 720  |
| aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat  | 780  |
| atgggaaact ccgtggtcac atacaacacc aacatcgacg aactgtgcaa gagcggcaag | 840  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg  | 900  |
| gtcacatata ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc | 960  |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtgac atacaccacc | 1020 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag | 1080 |
| atgaaggaga ggggcatcaa gcccgcgtg gtcacatata acaccctgat cgacggcctg  | 1140 |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc | 1200 |

|            |             |            |            |             |            |      |
|------------|-------------|------------|------------|-------------|------------|------|
| atcaccccca | gcgtggttac  | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 1260 |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gcccacgtg  | 1320 |
| gtcacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaggtgga  | tgaggccctg | 1380 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccg | acgtggttac  | atacaacacc | 1440 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag | atcgaggagg | ccctgaagct  | gttcaaggag | 1500 |
| atggaggaga | agggcatacac | ccccagcgtg | gtcacataca | ccacactgat  | cgacggactg | 1560 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc | gagctgttca | aagagatgcg  | gagcaagggc | 1620 |
| gtgaagccca | acgtggtgac  | atacaccacc | ctgatcgacg | gcctgtgcaa  | ggccggcaaa | 1680 |
| gtggacgagg | ccctggagct  | gttcgacgag | atgaaggaga | ggggcatcaa  | gcccgcgtg  | 1740 |
| gtcacataca | acaccctgat  | cgacggcctg | tgcaagagcg | gcaagatcga  | ggaggccctg | 1800 |
| aagctgttca | aggagatgga  | ggagaagggc | atcaccccca | gcgtggttac  | atacaccaca | 1860 |
| ctgatcgacg | gactgtgtaa  | agccggcgac | gtggacgaag | ccctcgagct  | gttcaaagag | 1920 |
| atgcggagca | agggcgtgaa  | gcccacgtg  | gtcacataca | ccaccctgat  | cgacggcctg | 1980 |
| tgcaaggccg | gcaaggtgga  | tgaggccctg | gagctgttcg | acgagatgaa  | ggagaggggc | 2040 |
| atcaagcccg | acgtggtcac  | atacaacacc | ctgatcgacg | gcctgtgcaa  | gagcggcaag | 2100 |
| atcgaggagg | ccctgaagct  | gttcaaggag | atggaggaga | agggcatacac | ccccagcgtg | 2160 |
| gtcacataca | ccacactgat  | cgacggactg | tgtaaagccg | gcgacgtgga  | cgaagccctc | 2220 |
| gagctgttca | aagagatgcg  | gagcaagggc | gtgaagccca | acgtggttac  | atacaccacc | 2280 |
| ctgatcgacg | gcctgtgcaa  | ggccggcaaa | gtggacgagg | ccctggagct  | gttcgacgag | 2340 |
| atgaaggaga | ggggcatcaa  | gcccgcgtg  | gtcacataca | acaccctgat  | cgacggcctg | 2400 |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 2460 |
| atcaccccca | gcgtggtgac  | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 2520 |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gcccacgtg  | 2580 |
| gtcacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaggtgga  | tgaggccctg | 2640 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccg | acgagctgac  | ctacaacacc | 2700 |
| ctgatcagcg | gcctgggcaa  | ggccggcaga | gccagagacc | ccccctgct   | cagtagcggg | 2760 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac   | 2820 |
| gataagatgg ccggacgcta gmagvskgtg vvdgdvngh ksvsgggdat ygktkcttgk    | 2880 |
| vwtvttttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtvnrkgd kdgnghkyny  | 2940 |
| nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm    | 3000 |
| dykkkkkrkvs hgsggsggggh mgnsvvtynt ndcksgkakk mkgtsvvtyt tdgckagdvd | 3060 |
| akmrskgvkn vvttyttgdck agkvdadmk r gkdvvtyntd gcksgkakkm kgtsvvtytt | 3120 |
| dgckagdvd kmrskgvknv vtyttgdcka gkvdadmkrg kdvtyntdg cksgkakkmk     | 3180 |
| gtsvvtytt d gckagdvdak mrskgvknv tyttgdckag kvdadmkrgk dvtyntdgc    | 3240 |
| ksgkakkmkg tsvvtytt d ckagdvdakm rskgvknv yttgdckagk vdadmkrgkd     | 3300 |
| vvtyntdgck sgkakkmkg t svvtytt dgc kagdvdakmr skgvknvty ttdgckagkv  | 3360 |
| dadmkrgkdv vtyntdgcks gkakkmkg tsvvtytt dgc agdvdakmr kgvknvtyt     | 3420 |
| tdgckagkvd admkrgkdy ntsggkagra rdvssgdykd hdgdykd hdd ykdddkmag    | 3480 |
| r                                                                   | 3481 |

<210> 21  
 <211> 3482  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2858)..(2858)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2914)..(2914)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (2934)..(2934)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2939)..(2939)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2941)..(2941)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2944)..(2944)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2956)..(2956)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2960)..(2960)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2969)..(2969)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2976)..(2976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2986)..(2986)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3031)..(3031)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3071)..(3071)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3099)..(3099)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3140)..(3140)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3168)..(3168)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3209)..(3209)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3237)..(3237)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3278)..(3278)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3306)..(3306)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3347)..(3347)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3375)..(3375)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3416)..(3416)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3442)..(3442)

<223> n is a, c, g, t or u

<400> 21

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccac cctggctgag  | 60   |
| ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc  | 120  |
| acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg  | 180  |
| cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac  | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc  | 300  |
| atctttttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac  | 360  |
| accctgggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg | 420  |
| gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag  | 480  |
| aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag  | 540  |
| ctcgccgacc actaccagca gaacaccccc atcgggcgac gccccgtgtc gctgcccga   | 600  |
| aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa ggcgcatcac  | 660  |
| atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac  | 720  |
| aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat   | 780  |
| atgggaaact ccgtgggtcac atacaacacc aacatcgaca aactgtgcaa gagcggcaag | 840  |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg   | 900  |
| gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 960  |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtgac atacaccacc | 1020 |
| ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag  | 1080 |

|            |             |            |            |             |            |      |
|------------|-------------|------------|------------|-------------|------------|------|
| atgaaggaga | ggggcatcaa  | gcccgcgctg | gtcacataca | acaccctgat  | cgacggcctg | 1140 |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 1200 |
| atcaccccca | gcgtgggttac | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 1260 |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gcccacgtg  | 1320 |
| gtcacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaggtgga  | tgaggccctg | 1380 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccc | acgtgggttac | atacaacacc | 1440 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag | atcgaggagg | ccctgaagct  | gttcaaggag | 1500 |
| atggaggaga | agggcataca  | ccccagcgtg | gtcacataca | ccacactgat  | cgacggactg | 1560 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc | gagctgttca | aagagatgcg  | gagcaagggc | 1620 |
| gtgaagccca | acgtgggtgac | atacaccacc | ctgatcgacg | gcctgtgcaa  | ggccggcaaa | 1680 |
| gtggacgagg | ccctggagct  | gttcgacgag | atgaaggaga | ggggcatcaa  | gcccgcgctg | 1740 |
| gtcacataca | acaccctgat  | cgacggcctg | tgcaagagcg | gcaagatcga  | ggaggccctg | 1800 |
| aagctgttca | aggagatgga  | ggagaagggc | atcaccccca | gcgtgggttac | atacaccaca | 1860 |
| ctgatcgacg | gactgtgtaa  | agccggcgac | gtggacgaag | ccctcgagct  | gttcaaagag | 1920 |
| atgcggagca | agggcgtgaa  | gcccacgtg  | gtcacataca | ccaccctgat  | cgacggcctg | 1980 |
| tgcaaggccg | gcaaggtgga  | tgaggccctg | gagctgttcg | acgagatgaa  | ggagaggggc | 2040 |
| atcaagcccc | acgtgggtcac | atacaacacc | ctgatcgacg | gcctgtgcaa  | gagcggcaag | 2100 |
| atcgaggagg | ccctgaagct  | gttcaaggag | atggaggaga | agggcataca  | ccccagcgtg | 2160 |
| gtcacataca | ccacactgat  | cgacggactg | tgtaaagccg | gcgacgtgga  | cgaagccctc | 2220 |
| gagctgttca | aagagatgcg  | gagcaagggc | gtgaagccca | acgtgggttac | atacaccacc | 2280 |
| ctgatcgacg | gcctgtgcaa  | ggccggcaaa | gtggacgagg | ccctggagct  | gttcgacgag | 2340 |
| atgaaggaga | ggggcatcaa  | gcccgcgctg | gtcacataca | acaccctgat  | cgacggcctg | 2400 |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 2460 |
| atcaccccca | gcgtgggtgac | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 2520 |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gcccacgtg  | 2580 |
| gtcacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaggtgga  | tgaggccctg | 2640 |

gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgagctgac ctacaacacc 2700  
 ctgatcagcg gcctgggcaa ggccggcaga gccagagacc cccccgtgct cagtagcggg 2760  
 gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac 2820  
 gataagatgg ccggacgcta gmagvskgtg vvdgdvngh ksvsgggdat ygktkcttgk 2880  
 vwtvtttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtvnrkgd kdgngkhkyny 2940  
 nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm 3000  
 dykkkkkrkvs hgsggsggggh mgnsvvtynt ndkcksgkak kmkgtsvvty ttdgckagdv 3060  
 dakmrskgvk nvvtytttdgc kagkvdadm rkdvvtynt dgcksgkakk mkgtsvvtyt 3120  
 tdgckagdvd akmrskgvkn vvytttdgck agkvdadmkr gkdvvtyntd gcksgkakkm 3180  
 kgtsvvtytt dgckagdvd kmrskgvknv vtyttgdcka gkvdadmkr gkdvvtyntdg 3240  
 cksgkakkmk gtsvvtytttd gckagdvdak mrskgvknv tyttgdckag kvdadmkr gk 3300  
 dvvtyntdgc ksgkakkmkg tsvvtytttdg ckagdvdakm rskgvknvvt yttgdckagk 3360  
 vdadmkr gkd vvyntdgc sggkakkmkg svvtytttdgc kagdvdakmr skgvknvty 3420  
 ttdgckagkv dadmkr gkdt yntsggkagr ardvssgdyk dhgdgdykdhd dykdddkma 3480  
 gr 3482

<210> 22  
 <211> 3482  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2858)..(2858)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2914)..(2914)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2934)..(2934)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2939)..(2939)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2941)..(2941)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2944)..(2944)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2952)..(2952)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2956)..(2956)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2960)..(2960)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2969)..(2969)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2976)..(2976)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2986)..(2986)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3029)..(3029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3071)..(3071)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3099)..(3099)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3140)..(3140)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3168)..(3168)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3209)..(3209)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3237)..(3237)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3278)..(3278)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3306)..(3306)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3347)..(3347)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3375)..(3375)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3416)..(3416)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3442)..(3442)  
 <223> n is a, c, g, t or u

<400> 22  
 atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccac cctggctcgag 60  
 ctggacggcg acgtaaaccg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc 120  
 acctacggca agctgaccct gaagtctatc tgcaccaccg gcaagctgcc cgtgccctgg 180  
 cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac 240  
 atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc 300  
 atctttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac 360  
 accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg 420  
 gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag 480  
 aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag 540  
 ctgcccgaac actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgaac 600  
 aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac 660  
 atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac 720  
 aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg agggctccat 780  
 atgggaaact ccgtgggtcac atacaacacc gatatcgacg gcctgtgcaa gagcggcaag 840  
 atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg 900  
 gtcacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc 960  
 gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtggtgac atacaccacc 1020  
 ctgatcgacg gcctgtgcaa ggccggcaaa gtggacgagg ccctggagct gttcgacgag 1080

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| atgaaggaga  | ggggcatcaa  | gcccgcgctg  | gtcacataca  | acaccctgat  | cgacggcctg  | 1140 |
| tgcaagagcg  | gcaagatcga  | ggaggccctg  | aagctgttca  | aggagatgga  | ggagaagggc  | 1200 |
| atcaccccca  | gcgtgggttac | atacaccaca  | ctgatcgacg  | gactgtgtaa  | agccggcgac  | 1260 |
| gtggacgaag  | ccctcgagct  | gttcaaagag  | atgcggagca  | agggcgtgaa  | gcccacgtg   | 1320 |
| gtcacataca  | ccaccctgat  | cgacggcctg  | tgcaaggccg  | gcaaggtgga  | tgaggccctg  | 1380 |
| gagctgttcg  | acgagatgaa  | ggagaggggc  | atcaagccccg | acgtgggttac | atacaacacc  | 1440 |
| ctgatcgacg  | gcctgtgcaa  | gagcggcaag  | atcgaggagg  | ccctgaagct  | gttcaaggag  | 1500 |
| atggaggaga  | agggcatac   | ccccagcgtg  | gtcacataca  | ccacactgat  | cgacggactg  | 1560 |
| tgtaaagccg  | gcgacgtgga  | cgaagccctc  | gagctgttca  | aagagatgcg  | gagcaagggc  | 1620 |
| gtgaagccca  | acgtgggtgac | atacaccacc  | ctgatcgacg  | gcctgtgcaa  | ggccggcaaaa | 1680 |
| gtggacgagg  | ccctggagct  | gttcgacgag  | atgaaggaga  | ggggcatcaa  | gcccgcgctg  | 1740 |
| gtcacataca  | acaccctgat  | cgacggcctg  | tgcaagagcg  | gcaagatcga  | ggaggccctg  | 1800 |
| aagctgttca  | aggagatgga  | ggagaagggc  | atcaccccca  | gcgtgggttac | atacaccaca  | 1860 |
| ctgatcgacg  | gactgtgtaa  | agccggcgac  | gtggacgaag  | ccctcgagct  | gttcaaagag  | 1920 |
| atgcggagca  | agggcgtgaa  | gcccacgtg   | gtcacataca  | ccaccctgat  | cgacggcctg  | 1980 |
| tgcaaggccg  | gcaaggtgga  | tgaggccctg  | gagctgttcg  | acgagatgaa  | ggagaggggc  | 2040 |
| atcaagccccg | acgtgggtcac | atacaacacc  | ctgatcgacg  | gcctgtgcaa  | gagcggcaag  | 2100 |
| atcgaggagg  | ccctgaagct  | gttcaaggag  | atggaggaga  | agggcatac   | ccccagcgtg  | 2160 |
| gtcacataca  | ccacactgat  | cgacggactg  | tgtaaagccg  | gcgacgtgga  | cgaagccctc  | 2220 |
| gagctgttca  | aagagatgcg  | gagcaagggc  | gtgaagccca  | acgtgggttac | atacaccacc  | 2280 |
| ctgatcgacg  | gcctgtgcaa  | ggccggcaaaa | gtggacgagg  | ccctggagct  | gttcgacgag  | 2340 |
| atgaaggaga  | ggggcatcaa  | gcccgcgctg  | gtcacataca  | acaccctgat  | cgacggcctg  | 2400 |
| tgcaagagcg  | gcaagatcga  | ggaggccctg  | aagctgttca  | aggagatgga  | ggagaagggc  | 2460 |
| atcaccccca  | gcgtgggtgac | atacaccaca  | ctgatcgacg  | gactgtgtaa  | agccggcgac  | 2520 |
| gtggacgaag  | ccctcgagct  | gttcaaagag  | atgcggagca  | agggcgtgaa  | gcccacgtg   | 2580 |
| gtcacataca  | ccaccctgat  | cgacggcctg  | tgcaaggccg  | gcaaggtgga  | tgaggccctg  | 2640 |
| gagctgttcg  | acgagatgaa  | ggagaggggc  | atcaagccccg | acgagctgac  | ctacaacacc  | 2700 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ctgatcagcg gcctgggcaa ggccggcaga gccagagacc cccccgtgct cagtagcggg   | 2760 |
| gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac   | 2820 |
| gataagatgg ccggacgcta gmagvskgtg vvdgdvngh ksvsgggdat ygktkcttgk    | 2880 |
| vwtvtttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtvnrkgd kdgnghkyny   | 2940 |
| nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyt saskdnkrdh mvvtaagtgm    | 3000 |
| dykkkkkrkvs hgsggsggggh mgnsvvtynt ddgcksgkak kmkgtsvvty ttdgckagdv | 3060 |
| dakmrskgvk nvvtyttgdc kagkvdadm rkgdvvtynt dgcksgkakk mkgtsvvtyt    | 3120 |
| tdgckagdv akmrskgvkn vvyttgdc agkvdadmkr gkdvvtyntd gcksgkakkm      | 3180 |
| kgtsvvtytt dgckagdvda kmrskgvknv vtyttgcka gkvdadmkr gkdvvtyntdg    | 3240 |
| cksgkakkm gtsvvtytt gckagdvda mrskgvknv tyttgckag kvdadmkr gk       | 3300 |
| dvvyntdgc ksgkakkmkg tsvvtyttg ckagdvda mrskgvknv tyttgckag         | 3360 |
| vdadmkr gkd vvtyntdgc sgkakkmkg svvtyttgdc kagdvda mr skgvknvty     | 3420 |
| ttdgckagkv dadmkr gkdt yntsggkagr ardvssgdyk dhgdgdykdhd dykdddddma | 3480 |
| gr                                                                  | 3482 |

<210> 23  
 <211> 3551  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2912)..(2912)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2982)..(2982)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2990)..(2990)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2996)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3007)..(3007)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3011)..(3011)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3015)..(3015)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3025)..(3025)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3032)..(3032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3042)..(3042)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3066)..(3066)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3072)..(3072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (3113)..(3113)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3141)..(3141)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3182)..(3182)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3210)..(3210)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3251)..(3251)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3279)..(3279)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3320)..(3320)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3348)..(3348)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3389)..(3389)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3417)..(3417)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3458)..(3458)  
 <223> n is a, c, g, t or u

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (3484)..(3484)

&lt;223&gt; n is a, c, g, t or u

&lt;400&gt; 23

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggccggag tgtccaaagg cgaggagctg ttaccggcg tcgtgcctat tctggtggag   | 60   |
| ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc  | 120  |
| acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg  | 180  |
| cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat   | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccagg gatacgtgca agaaaggacc   | 300  |
| atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac  | 360  |
| acactggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg  | 420  |
| ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag  | 480  |
| aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag  | 540  |
| ctggccgatc actaccagca gaacacaccc atcggcgatg gaccggtttt actgcccgcac | 600  |
| aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat  | 660  |
| atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac   | 720  |
| aagagccggc tccatatggg aaactccgtg gtcacataca acaccctgat cgacggcctg  | 780  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 840  |
| atcccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac  | 900  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg   | 960  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 1020 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtgggtcac atacaacacc | 1080 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 1140 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg   | 1200 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1260 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag | 1320 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg   | 1380 |
| gttacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1440 |

|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| aagctgttca | aggagatgga  | ggagaagggc  | atcaccccca | gcgtgggtcac | atacaccaca  | 1500 |
| ctgatcgacg | gactgtgtaa  | agccggcgac  | gtggacgaag | ccctcgagct  | gttcaaagag  | 1560 |
| atgcggagca | agggcgtgaa  | gccccaacgtg | gtgacataca | ccaccctgat  | cgacggcctg  | 1620 |
| tgcaaggccg | gcaaagtgga  | cgaggccctg  | gagctgttcg | acgagatgaa  | ggagaggggc  | 1680 |
| atcaagcccg | acgtgggtcac | atacaacacc  | ctgatcgacg | gcctgtgcaa  | gagcggcaag  | 1740 |
| atcgaggagg | ccctgaagct  | gttcaaggag  | atggaggaga | agggcacac   | ccccagcgtg  | 1800 |
| gttacataca | ccacactgat  | cgacggactg  | tgtaaagccg | gcgacgtgga  | cgaagccctc  | 1860 |
| gagctgttca | aagagatgcg  | gagcaagggc  | gtgaagccca | acgtgggtcac | atacaccacc  | 1920 |
| ctgatcgacg | gcctgtgcaa  | ggccggcaag  | gtggatgagg | ccctggagct  | gttcgacgag  | 1980 |
| atgaaggaga | ggggcatcaa  | gcccgcacgtg | gtcacataca | acaccctgat  | cgacggcctg  | 2040 |
| tgcaagagcg | gcaagatcga  | ggaggccctg  | aagctgttca | aggagatgga  | ggagaagggc  | 2100 |
| atcaccccca | gcgtgggtcac | atacaccaca  | ctgatcgacg | gactgtgtaa  | agccggcgac  | 2160 |
| gtggacgaag | ccctcgagct  | gttcaaagag  | atgcggagca | agggcgtgaa  | gccccaacgtg | 2220 |
| gttacataca | ccaccctgat  | cgacggcctg  | tgcaaggccg | gcaaagtgga  | cgaggccctg  | 2280 |
| gagctgttcg | acgagatgaa  | ggagaggggc  | atcaagcccg | acgtgggtcac | atacaacacc  | 2340 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag  | atcgaggagg | ccctgaagct  | gttcaaggag  | 2400 |
| atggaggaga | agggcacac   | ccccagcgtg  | gtgacataca | ccacactgat  | cgacggactg  | 2460 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc  | gagctgttca | aagagatgcg  | gagcaagggc  | 2520 |
| gtgaagccca | acgtgggtcac | atacaccacc  | ctgatcgacg | gcctgtgcaa  | ggccggcaag  | 2580 |
| gtggatgagg | ccctggagct  | gttcgacgag  | atgaaggaga | ggggcatcaa  | gcccgcagag  | 2640 |
| ctgacctaca | acaccctgat  | cagcggcctg  | ggcaaggccg | gcagagccag  | agaccccccc  | 2700 |
| gtgctcagta | gccccaaaga  | gaaacgcaaa  | gtcgaggatc | caaagaagaa  | aaggaaggtt  | 2760 |
| gaagacccca | agaaaaagag  | gaagggtggg  | tccgactata | aggaccacga  | cggagactac  | 2820 |
| aaggatcatg | atattgatta  | caaagacgat  | gacgataaga | tggccccaaa  | gaagaagcgg  | 2880 |
| aaggtcggac | gctagmagvs  | kgtgvvvdgd  | vnghksvrgg | gdatngktkc  | ttgkvwtvtt  | 2940 |
| gygvacsryd | hmkhdksamg  | yvrtskddgt  | yktravkgdt | vnrkgdkdgn  | ghkynnshyv  | 3000 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ytadkkncka nkrhnvdgsv adhyntgdgv dnhyshsksk dnkrdhmvvt aagthgmdyk   | 3060 |
| srhmgnsvvt yntdgcksgk akkmkgtsvv tyttgdckag dvdakmrskg vknvtytttd   | 3120 |
| gckagkvdad mkrkgdvvtv ntdgcksgka kkmkgtsvvv yttdgckagd vdakmrskgv   | 3180 |
| knvtytttdg ckagkvdadm krgkdvvtyn tdgcksgkak kmkgtsvvty ttdgckagdv   | 3240 |
| dakmrskgvk nvvtytttdgc kagkvdadmk rgkdvvtynt dgcksgkak mkgtsvvtyt   | 3300 |
| tdgckagdvda akmrskgvkn vvttyttgdck agkvdadmkr gkdvvtyntd gcksgkakkm | 3360 |
| kgtsvvtytt dgckagdvda kmrskgvknv vtyttgdcka gkvdadmkr gkdvvtyntd    | 3420 |
| cksgkakkmk gtsvvtytttd gckagdvda mrskgvknv tyttgdckag kvdadmkr gk   | 3480 |
| dyntsggka grandvsskk krkvdkkkkrk vdkkkkrkvg dykdhdgdyk dhddykdddd   | 3540 |
| kmakkkkrkv r                                                        | 3551 |

<210> 24  
 <211> 3550  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2912)..(2912)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2982)..(2982)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2990)..(2990)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2996)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3007)..(3007)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3011)..(3011)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3015)..(3015)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3025)..(3025)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3032)..(3032)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3042)..(3042)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3066)..(3066)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3072)..(3072)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3112)..(3112)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3140)..(3140)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (3181)..(3181)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3209)..(3209)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3250)..(3250)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3278)..(3278)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3319)..(3319)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3347)..(3347)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3388)..(3388)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3416)..(3416)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3457)..(3457)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3483)..(3483)  
 <223> n is a, c, g, t or u

<400> 24  
 atggccggag tgtccaaagg cgaggagctg tttaccggcg tcgtgcctat tctggtggag 60  
 ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc 120

|                                                                    |      |
|--------------------------------------------------------------------|------|
| acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg  | 180  |
| cccacactgg tgaccacctt cggctacggc gtggcttgtt tctctcgta ccccgaccat   | 240  |
| atgaagcagc acgacttctt caagtccgcc atgcccagg gatacgtgca agaaaggacc   | 300  |
| atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac  | 360  |
| acactgggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg | 420  |
| ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag  | 480  |
| aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag  | 540  |
| ctggccgatc actaccagca gaacacaccc atcggcgatg gaccgtttt actgcccgac   | 600  |
| aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat  | 660  |
| atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac   | 720  |
| aagagccggc tccatatggg aaactccgtg gtcacataca acaccctgat cgacgaactg  | 780  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 840  |
| atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac | 900  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gccaacgtg   | 960  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 1020 |
| gagctgttcg acgagatgaa ggagaggggc atcaagccc acgtgggtcac atacaacacc  | 1080 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 1140 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg   | 1200 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1260 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag | 1320 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgtg   | 1380 |
| gttacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1440 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca | 1500 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag  | 1560 |
| atgcggagca agggcgtgaa gccaacgtg gtgacataca ccaccctgat cgacggcctg   | 1620 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc  | 1680 |

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| atcaagcccg | acgtggtcac | atacaacacc | ctgatcgacg | gcctgtgcaa | gagcggcaag | 1740 |
| atcgaggagg | ccctgaagct | gttcaaggag | atggaggaga | agggcacac  | ccccagcgtg | 1800 |
| gttacatata | ccacactgat | cgacggactg | tgtaaagccg | gcgacgtgga | cgaagccctc | 1860 |
| gagctgttca | aagagatgcg | gagcaagggc | gtgaagccca | acgtggtcac | atacaccacc | 1920 |
| ctgatcgacg | gcctgtgcaa | ggccggcaag | gtggatgagg | ccctggagct | gttcgacgag | 1980 |
| atgaaggaga | ggggcatcaa | gcccgcgctg | gtcacatata | acaccctgat | cgacggcctg | 2040 |
| tgcaagagcg | gcaagatcga | ggaggccctg | aagctgttca | aggagatgga | ggagaagggc | 2100 |
| atcaccccca | gcgtggtcac | atacaccaca | ctgatcgacg | gactgtgtaa | agccggcgac | 2160 |
| gtggacgaag | ccctcgagct | gttcaaagag | atgcggagca | agggcgtgaa | gcccacgtg  | 2220 |
| gttacatata | ccaccctgat | cgacggcctg | tgcaaggccg | gcaaagtgga | cgaggccctg | 2280 |
| gagctgttcg | acgagatgaa | ggagaggggc | atcaagcccg | acgtggtcac | atacaacacc | 2340 |
| ctgatcgacg | gcctgtgcaa | gagcggcaag | atcgaggagg | ccctgaagct | gttcaaggag | 2400 |
| atggaggaga | agggcacac  | ccccagcgtg | gtgacatata | ccacactgat | cgacggactg | 2460 |
| tgtaaagccg | gcgacgtgga | cgaagccctc | gagctgttca | aagagatgcg | gagcaagggc | 2520 |
| gtgaagccca | acgtggtcac | atacaccacc | ctgatcgacg | gcctgtgcaa | ggccggcaag | 2580 |
| gtggatgagg | ccctggagct | gttcgacgag | atgaaggaga | ggggcatcaa | gcccgcgag  | 2640 |
| ctgacctata | acaccctgat | cagcggcctg | ggcaaggccg | gcagagccag | agaccccccc | 2700 |
| gtgctcagta | gccccaaaga | gaaacgcaaa | gtcgaggatc | caaagaagaa | aaggaaggtt | 2760 |
| gaagacccca | agaaaaagag | gaaggtgggt | tccgactata | aggaccacga | cggagactac | 2820 |
| aaggatcatg | atattgatta | caaagacgat | gacgataaga | tggccccaaa | gaagaagcgg | 2880 |
| aaggtcggac | gctagmagvs | kgtgvvvdgd | vnghksvrgg | gdatngktkc | ttgkvwtvtt | 2940 |
| gygvacsryd | hmkhdksamg | yvrtskddgt | yktravkgdt | vnrkgdkdgn | ghkynnshyv | 3000 |
| ytadkkncka | nkrhnvdgsv | adhyntgdgv | dnhyshsksk | dnkrdhmvvt | aagthgmdyk | 3060 |
| srhmgnsvvt | yntdcksgka | kkmkgtsvvt | yttgdckagd | vdakmrskgv | knvvtyttgd | 3120 |
| ckagkvdadm | krqkdvvtyn | tdgcksgkak | kmkgtsvvty | ttgdckagdv | dakmrskgvk | 3180 |
| nvvttyttgd | cagkvdadmk | rgkdvvtynt | dgcksgkakk | mkgtsvvtyt | tdgckagdv  | 3240 |
| akmrskgvkn | vvtyttgdck | agkvdadmkr | gkdvvtyntd | gcksgkakkm | kgtsvvtytt | 3300 |

```

dgckagdvda kmrskgvknv vtyttgdgcka gkvdadmkrk kdvvtyntdg cksgkakkmk      3360
gtsvtytttd gckagdvda kmrskgvknv tyttgdgckag kvdadmkrk dvvtyntdgc      3420
ksgkakkmk gtsvtytttdg ckagdvda kmrskgvknv tyttgdgckag kvdadmkrkd      3480
tyntsggkag rardvsskkk rkvdkkkrkv dkkkrkvgsd ykdhgdgdyk hddykdddk      3540
makkkrvgr                                                                    3550

```

```

<210> 25
<211> 3551
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> PPR vector

```

```

<220>
<221> misc_feature
<222> (2912)..(2912)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2925)..(2925)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2982)..(2982)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2990)..(2990)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2995)..(2996)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (3007)..(3007)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature

```

<222> (3011)..(3011)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3015)..(3015)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3025)..(3025)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3032)..(3032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3042)..(3042)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3066)..(3066)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3072)..(3072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3074)..(3074)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3113)..(3113)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3141)..(3141)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3182)..(3182)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3210)..(3210)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3251)..(3251)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3279)..(3279)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3320)..(3320)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3348)..(3348)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3389)..(3389)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3417)..(3417)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3458)..(3458)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3484)..(3484)  
 <223> n is a, c, g, t or u

<400> 25  
 atggccggag tgtccaaagg cgaggagctg tttaccggcg tcgtgcctat tctggtggag 60  
 ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc 120  
 acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg 180  
 cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat 240

|            |             |            |            |             |            |      |
|------------|-------------|------------|------------|-------------|------------|------|
| atgaagcagc | acgacttctt  | caagtccgcc | atgcccagg  | gatacgtgca  | agaaaggacc | 300  |
| atctccttca | aggacgatgg  | cacctacaag | accagagccg | aggtgaagtt  | cgagggcgac | 360  |
| acactggtga | atcgtatcga  | actgaagggc | atcgacttca | aagaggacgg  | caacattctg | 420  |
| ggccacaagc | tggagtacaa  | cttcaacagc | cactacgtgt | acatcaccgc  | cgataagcag | 480  |
| aagaactgca | tcaaggccaa  | cttcaagatt | cgtcacaacg | tggaggatgg  | ctccgtgcag | 540  |
| ctggccgatc | actaccagca  | gaacacaccc | atcggcgatg | gacccgtttt  | actgcccgcg | 600  |
| aaccactatt | taagccacca  | gagcaagctg | tccaaggacc | ccaacgagaa  | gcgtgatcat | 660  |
| atggtgctgc | tcgagtttgt  | gaccgccgcc | ggcatcacc  | atggaatgga  | cgagctgtac | 720  |
| aagagccggc | tccatatggg  | aaactccgtg | gtcacataca | acaccaacat  | cgaccagctg | 780  |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 840  |
| atcaccccca | gcgtgggtcac | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 900  |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gcccacgtg  | 960  |
| gtgacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaagtgga  | cgaggccctg | 1020 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccg | acgtgggtcac | atacaacacc | 1080 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag | atcgaggagg | ccctgaagct  | gttcaaggag | 1140 |
| atggaggaga | agggcacac   | ccccagcgtg | gttacataca | ccacactgat  | cgacggactg | 1200 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc | gagctgttca | aagagatgcg  | gagcaagggc | 1260 |
| gtgaagccca | acgtgggtcac | atacaccacc | ctgatcgacg | gcctgtgcaa  | ggccggcaag | 1320 |
| gtggatgagg | ccctggagct  | gttcgacgag | atgaaggaga | ggggcatcaa  | gcccgcgtg  | 1380 |
| gttacataca | acaccctgat  | cgacggcctg | tgcaagagcg | gcaagatcga  | ggaggccctg | 1440 |
| aagctgttca | aggagatgga  | ggagaagggc | atcaccccca | gcgtgggtcac | atacaccaca | 1500 |
| ctgatcgacg | gactgtgtaa  | agccggcgac | gtggacgaag | ccctcgagct  | gttcaaagag | 1560 |
| atgcggagca | agggcgtgaa  | gcccacgtg  | gtgacataca | ccaccctgat  | cgacggcctg | 1620 |
| tgcaaggccg | gcaaagtgga  | cgaggccctg | gagctgttcg | acgagatgaa  | ggagaggggc | 1680 |
| atcaagcccg | acgtgggtcac | atacaacacc | ctgatcgacg | gcctgtgcaa  | gagcggcaag | 1740 |
| atcgaggagg | ccctgaagct  | gttcaaggag | atggaggaga | agggcacac   | ccccagcgtg | 1800 |

|            |             |             |             |             |            |      |
|------------|-------------|-------------|-------------|-------------|------------|------|
| gttacataca | ccacactgat  | cgacggactg  | tgtaaagccg  | gcgacgtgga  | cgaagccctc | 1860 |
| gagctgttca | aagagatgcg  | gagcaagggc  | gtgaagccca  | acgtgggtcac | atacaccacc | 1920 |
| ctgatcgacg | gcctgtgcaa  | ggccggcaag  | gtggatgagg  | ccctggagct  | gttcgacgag | 1980 |
| atgaaggaga | ggggcatcaa  | gcccgcgctg  | gtcacataca  | acaccctgat  | cgacggcctg | 2040 |
| tgcaagagcg | gcaagatcga  | ggaggccctg  | aagctgttca  | aggagatgga  | ggagaagggc | 2100 |
| atcaccccca | gcgtgggtcac | atacaccaca  | ctgatcgacg  | gactgtgtaa  | agccggcgac | 2160 |
| gtggacgaag | ccctcgagct  | gttcaaagag  | atgcggagca  | agggcgtgaa  | gccaacgtg  | 2220 |
| gttacataca | ccaccctgat  | cgacggcctg  | tgcaaggccg  | gcaaagtgga  | cgaggccctg | 2280 |
| gagctgttcg | acgagatgaa  | ggagaggggc  | atcaagcccg  | acgtgggtcac | atacaacacc | 2340 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag  | atcgaggagg  | ccctgaagct  | gttcaaggag | 2400 |
| atggaggaga | agggcatacac | ccccagcgtg  | gtgacataca  | ccacactgat  | cgacggactg | 2460 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc  | gagctgttca  | aagagatgcg  | gagcaagggc | 2520 |
| gtgaagccca | acgtgggtcac | atacaccacc  | ctgatcgacg  | gcctgtgcaa  | ggccggcaag | 2580 |
| gtggatgagg | ccctggagct  | gttcgacgag  | atgaaggaga  | ggggcatcaa  | gcccgcgacg | 2640 |
| ctgacctaca | acaccctgat  | cagcggcctg  | ggcaaggccg  | gcagagccag  | agaccccccc | 2700 |
| gtgctcagta | gccccaaaga  | gaaacgcaaa  | gtcgaggatc  | caaagaagaa  | aaggaaggtt | 2760 |
| gaagacccca | agaaaaagag  | gaaggtgggt  | tccgactata  | aggaccacga  | cggagactac | 2820 |
| aaggatcatg | atattgatta  | caaagacgat  | gacgataaga  | tggccccaaa  | gaagaagcgg | 2880 |
| aaggtcggac | gctagmagvs  | kgtgvvvdgd  | vnhksvrgg   | gdatngktkc  | ttgkvwtvtt | 2940 |
| gygvacsryd | hmkhdksamg  | yvrtskddgt  | yktravkgdt  | vnrkgdkdgn  | ghkynnshyv | 3000 |
| ytadkkncka | nkrhndgsv   | adhyntgdgv  | dnhyshsksk  | dnkrdhmvvt  | aagthgmdyk | 3060 |
| srhmgnsvvt | yntndcksgk  | akkmkgtsvv  | tyttgdgckag | dvdakmrskg  | vknvtytttd | 3120 |
| gckagkvdad | mkrkgdvvty  | ntdgcksgka  | kkmkgtsvvt  | yttdgckagd  | vdakmrskgv | 3180 |
| knvtytttdg | ckagkvdadm  | krkgdvvtyn  | tdgcksgkak  | kkmkgtsvty  | ttdgckagdv | 3240 |
| dakmrskgvk | nvtytttdgc  | kagkvdadmk  | rgkdvvtynt  | dgcksgkakk  | mkgtsvvtyt | 3300 |
| tdgckagdv  | akmrskgvkn  | vvtytttdgck | agkvdadmk   | gkdvvtyntd  | gcksgkakkm | 3360 |
| kgtsvvtytt | dgckagdvda  | kmrskgvknv  | vtytttdgcka | gkvdadmkrg  | kdvvtyntdg | 3420 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| cksgkakkmk gtsvvtyttdd gckagdvdak mrskgvknvv tyttgdgckag kvdadmkrgk | 3480 |
| dtyntsggka grandvsskk krkvdkkkkrk vdkkkkrkvgs dykdhdgdyk dhddykdddd | 3540 |
| kmakkkkrkv g r                                                      | 3551 |

<210> 26  
 <211> 3551  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2912)..(2912)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2982)..(2982)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2990)..(2990)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2996)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3007)..(3007)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3011)..(3011)  
 <223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (3015)..(3015)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3025)..(3025)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3032)..(3032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3042)..(3042)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3066)..(3066)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3072)..(3072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3074)..(3074)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3113)..(3113)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3141)..(3141)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3182)..(3182)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3210)..(3210)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3251)..(3251)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3279)..(3279)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3320)..(3320)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3348)..(3348)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3389)..(3389)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3417)..(3417)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3458)..(3458)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3484)..(3484)  
 <223> n is a, c, g, t or u

<400> 26  
 atggccggag tgtccaaagg cgaggagctg tttaccggcg tcgtgcctat tctggtggag 60  
 ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc 120  
 aaaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg 180  
 cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat 240  
 atgaagcagc acgacttctt caagtccgcc atgcccagg gatacgtgca agaaaggacc 300  
 atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac 360

|                                                                    |      |
|--------------------------------------------------------------------|------|
| acactggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg  | 420  |
| ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag  | 480  |
| aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag  | 540  |
| ctggccgatc actaccagca gaacacaccc atcggcgatg gacccgtttt actgcccgac  | 600  |
| aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat  | 660  |
| atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac   | 720  |
| aagagccggc tccatatggg aaactccgtg gtcacataca acaccaacat cgacgaactg  | 780  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 840  |
| atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac | 900  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gccaacgtg   | 960  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 1020 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtgggtcac atacaacacc | 1080 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 1140 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg   | 1200 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 1260 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag | 1320 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgacgtg  | 1380 |
| gttacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg  | 1440 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca | 1500 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag  | 1560 |
| atgcggagca agggcgtgaa gccaacgtg gtgacataca ccaccctgat cgacggcctg   | 1620 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc  | 1680 |
| atcaagcccg acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag | 1740 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg   | 1800 |
| gttacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc  | 1860 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc | 1920 |

|              |             |             |             |             |              |      |
|--------------|-------------|-------------|-------------|-------------|--------------|------|
| ctgatcgacg   | gcctgtgcaa  | ggccggcaag  | gtggatgagg  | ccctggagct  | gttcgacgag   | 1980 |
| atgaaggaga   | ggggcatcaa  | gcccgcgctg  | gtcacataca  | acaccctgat  | cgacggcctg   | 2040 |
| tgcaagagcg   | gcaagatcga  | ggaggccctg  | aagctgttca  | aggagatgga  | ggagaagggc   | 2100 |
| atcaccccca   | gcgtgggtcac | atacaccaca  | ctgatcgacg  | gactgtgtaa  | agccggcgac   | 2160 |
| gtggacgaag   | ccctcgagct  | gttcaaagag  | atgcggagca  | agggcgtgaa  | gcccacgtg    | 2220 |
| gttacataca   | ccaccctgat  | cgacggcctg  | tgcaaggccg  | gcaaagtgga  | cgaggccctg   | 2280 |
| gagctgttcg   | acgagatgaa  | ggagagggggc | atcaagcccg  | acgtgggtcac | atacaacacc   | 2340 |
| ctgatcgacg   | gcctgtgcaa  | gagcggcaag  | atcgaggagg  | ccctgaagct  | gttcaaggag   | 2400 |
| atggaggaga   | agggcacac   | ccccagcgtg  | gtgacataca  | ccacactgat  | cgacggactg   | 2460 |
| tgtaaagccg   | gcgacgtgga  | cgaagccctc  | gagctgttca  | aagagatgcg  | gagcaagggc   | 2520 |
| gtgaagccca   | acgtgggtcac | atacaccacc  | ctgatcgacg  | gcctgtgcaa  | ggccggcaag   | 2580 |
| gtggatgagg   | ccctggagct  | gttcgacgag  | atgaaggaga  | ggggcatcaa  | gcccgcgag    | 2640 |
| ctgacctaca   | acaccctgat  | cagcggcctg  | ggcaaggccg  | gcagagccag  | agaccccccc   | 2700 |
| gtgctcagta   | gccccaaaga  | gaaacgcaaa  | gtcgaggatc  | caaagaagaa  | aaggaagggtt  | 2760 |
| gaagacccca   | agaaaaagag  | gaagggtgggt | tccgactata  | aggaccacga  | cggagactac   | 2820 |
| aaggatcatg   | atattgatta  | caaagacgat  | gacgataaga  | tggccccaaa  | gaagaagcgg   | 2880 |
| aaggtcggac   | gctagmagvs  | kgtgvvvdgd  | vnghksvrgg  | gdatngktkc  | ttgkvwtvtt   | 2940 |
| gygvacsryd   | hmkhdksamg  | yvrtskddgt  | yktravkgdt  | vnrkgdkgdn  | ghkynnshyv   | 3000 |
| ytadkkncka   | nkrhnvdgsv  | adhyntgdgv  | dnhyshsksk  | dnkrdhmvvt  | aagthgmdyk   | 3060 |
| srhmgnsvvt   | yntndcksgk  | akkmkgtsvv  | tytt dgckag | dvdakmrskg  | vknvtytttd   | 3120 |
| gckagkvdad   | mkr gkdvvty | ntdgcksgka  | kkmkgtsvvt  | ytt dgckagd | vdakmrskgv   | 3180 |
| knvvttytt dg | ckagkvdadm  | kr gkdvvty  | tdgcksgkak  | kkmkgtsvvt  | tytt dgckagd | 3240 |
| dakmrskgvk   | nvvttytt dg | kagkvdadmk  | rgkdvvtynt  | dgcksgkakk  | mkgtsvvtyt   | 3300 |
| tdgckagdv    | akmrskgvkn  | vvtytt dgck | agkvdadmk   | gkdvvtyntd  | gcksgkakkm   | 3360 |
| kgtsvvttytt  | dgckagdvda  | kmrskgvknv  | vytt dgcka  | gkvdadmkrg  | kdvvtyn t dg | 3420 |
| cksgkakkmk   | gtsvvttytt  | gckagdvda   | mrskgvknv   | tytt dgckag | kvdadmkr gk  | 3480 |
| dyntsggka    | grandvsskk  | krkvdkkkkrk | vdkkkrkvgs  | dykdhdgdyk  | dhddykdddd   | 3540 |

kmakkkkrkvg r

3551

<210> 27  
<211> 3552  
<212> DNA  
<213> Artificial sequence

<220>  
<223> PPR vector

<220>  
<221> misc\_feature  
<222> (2912)..(2912)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2925)..(2925)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2982)..(2982)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2990)..(2990)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2995)..(2996)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (3007)..(3007)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (3011)..(3011)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (3015)..(3015)  
<223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3025)..(3025)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3032)..(3032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3042)..(3042)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3066)..(3066)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3072)..(3072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3074)..(3074)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3114)..(3114)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3142)..(3142)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3183)..(3183)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3211)..(3211)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3252)..(3252)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3280)..(3280)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3321)..(3321)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3349)..(3349)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3390)..(3390)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3418)..(3418)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3459)..(3459)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3485)..(3485)

<223> n is a, c, g, t or u

<400> 27

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| atggccggag tgtccaaagg cgaggagctg ttaccggcg tcgtgcctat tctggtggag  | 60  |
| ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc | 120 |
| acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg | 180 |
| cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat  | 240 |
| atgaagcagc acgacttctt caagtccgcc atgcccagg gatactgca agaaaggacc   | 300 |
| atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac | 360 |
| acactggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg | 420 |
| ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag | 480 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag   | 540  |
| ctggccgatc actaccagca gaacacacccc atcggcgatg gacccgtttt actgcccgcac | 600  |
| aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat   | 660  |
| atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac    | 720  |
| aagagccggc tccatatggg aaactccgtg gtcacataca acaccaacat cgacaaactg   | 780  |
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc   | 840  |
| atcaccccca gcgtgggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac  | 900  |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gccaacgtg    | 960  |
| gtgacataca ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg   | 1020 |
| gagctgttcg acgagatgaa ggagaggggc atcaagccc acgtgggtcac atacaacacc   | 1080 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag   | 1140 |
| atggaggaga agggcatcac cccagcgtg gttacataca ccacactgat cgacggactg    | 1200 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc   | 1260 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag  | 1320 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgacgtg   | 1380 |
| gttacataca acaccctgat cgacggcctg tgcaagagcg gcaagatcga ggaggccctg   | 1440 |
| aagctgttca aggagatgga ggagaagggc atcaccccca gcgtgggtcac atacaccaca  | 1500 |
| ctgatcgacg gactgtgtaa agccggcgac gtggacgaag ccctcgagct gttcaaagag   | 1560 |
| atgcggagca agggcgtgaa gccaacgtg gtgacataca ccaccctgat cgacggcctg    | 1620 |
| tgcaaggccg gcaaagtgga cgaggccctg gagctgttcg acgagatgaa ggagaggggc   | 1680 |
| atcaagccc acgtgggtcac atacaacacc ctgatcgacg gcctgtgcaa gagcggcaag   | 1740 |
| atcgaggagg ccctgaagct gttcaaggag atggaggaga agggcatcac cccagcgtg    | 1800 |
| gttacataca ccacactgat cgacggactg tgtaaagccg gcgacgtgga cgaagccctc   | 1860 |
| gagctgttca aagagatgcg gagcaagggc gtgaagccca acgtgggtcac atacaccacc  | 1920 |
| ctgatcgacg gcctgtgcaa ggccggcaag gtggatgagg ccctggagct gttcgacgag   | 1980 |
| atgaaggaga ggggcatcaa gcccgacgtg gtcacataca acaccctgat cgacggcctg   | 2040 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tgcaagagcg gcaagatcga ggaggccctg aagctgttca aggagatgga ggagaagggc  | 2100 |
| atcaccccca gcgtggtcac atacaccaca ctgatcgacg gactgtgtaa agccggcgac  | 2160 |
| gtggacgaag ccctcgagct gttcaaagag atgcggagca agggcgtgaa gcccacgtg   | 2220 |
| gttacatata ccaccctgat cgacggcctg tgcaaggccg gcaaagtgga cgaggccctg  | 2280 |
| gagctgttcg acgagatgaa ggagaggggc atcaagcccg acgtggtcac atacaacacc  | 2340 |
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag  | 2400 |
| atggaggaga agggcatcac ccccgagcgtg gtgacatata ccacactgat cgacggactg | 2460 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc  | 2520 |
| gtgaagccca acgtggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag  | 2580 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcgag   | 2640 |
| ctgacctata acaccctgat cagcggcctg ggcaaggccg gcagagccag agaccccccc  | 2700 |
| gtgctcagta gcccgaagaa gaaacgcaaa gtcgaggatc caaagaagaa aaggaaggtt  | 2760 |
| gaagacccca agaaaaagag gaaggtgggt tccgactata aggaccacga cggagactac  | 2820 |
| aaggatcatg atattgatta caaagacgat gacgataaga tggcccaaaa gaagaagcgg  | 2880 |
| aaggtcggac gctagmagvs kgtgvvvdgd vnghksvrgg gdatngktkc ttgkvwtvtt  | 2940 |
| gygvacsryd hmkhdksamg yvrtskddgt yktravkgdt vnrkgdkdgn ghkynnshyv  | 3000 |
| ytadkkncka nkrhnvdgsv adhyntgdgv dnhyshsksk dnkrdhmvvt aagthgmdyk  | 3060 |
| srhmgnsvvt yntndkcksg kakkmkgtsv vtyttgcka gdvdakmrsk gvknvtytt    | 3120 |
| dgckagkvda dmkrkgdvvt yntdgcksgk akkmkgtsvv tyttgdckag dvdakmrskg  | 3180 |
| vknvtytttd gckagkvdad mkrkgdvvt yntdgcksgka kkmkgtsvvt yttgdckagd  | 3240 |
| vdakmrskgv knvtytttdg ckagkvdadm krgkdvvtyn tdgcksgkak kmkgtsvvt y | 3300 |
| ttgdckagdv dakmrskgvk nvvtytttdg kagkvdadmk rgkdvvtynt dgcksgkakk  | 3360 |
| mkgtsvvtyt tdgckagdvd akmrskgvkn vvtyttgdck agkvdadmkr gkdvvtyntd  | 3420 |
| gcksgkakkm kgtsvvtytt dgckagdvd kmrskgvknv vtyttgdcka gkvdadmkrg   | 3480 |
| kdyntsggk agrandvssk kkrkvdkkkv kvdkkkkrkv sdykdhdgdy kdhdyykddd   | 3540 |
| dkmakkkrv gr                                                       | 3552 |

<210> 28  
 <211> 3552  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> PPR vector

<220>  
 <221> misc\_feature  
 <222> (2912)..(2912)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2982)..(2982)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2990)..(2990)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2996)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3007)..(3007)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3011)..(3011)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3015)..(3015)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3025)..(3025)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3032)..(3032)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3042)..(3042)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3066)..(3066)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3072)..(3072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3114)..(3114)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3142)..(3142)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3183)..(3183)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3211)..(3211)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3252)..(3252)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3280)..(3280)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (3321)..(3321)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3349)..(3349)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3390)..(3390)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3418)..(3418)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3459)..(3459)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3485)..(3485)  
 <223> n is a, c, g, t or u

<400> 28  
 atggccggag tgtccaaagg cgaggagctg tttaccggcg tcgtgcctat tctggtggag 60  
 ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc 120  
 acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg 180  
 cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat 240  
 atgaagcagc acgacttctt caagtccgcc atgcccagg gatacgtgca agaaaggacc 300  
 atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac 360  
 aactggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg 420  
 ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag 480  
 aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag 540  
 ctggccgatc actaccagca gaacacaccc atcggcgatg gacccgtttt actgcccgac 600  
 aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat 660  
 atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac 720

|            |             |            |            |             |            |      |
|------------|-------------|------------|------------|-------------|------------|------|
| aagagccggc | tccatatggg  | aaactccgtg | gtcacataca | acaccgatat  | cgacggcctg | 780  |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 840  |
| atcaccccca | gcgtgggtcac | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 900  |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gccccacgtg | 960  |
| gtgacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaagtgga  | cgaggccctg | 1020 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccc | acgtgggtcac | atacaacacc | 1080 |
| ctgatcgacg | gcctgtgcaa  | gagcggcaag | atcgaggagg | ccctgaagct  | gttcaaggag | 1140 |
| atggaggaga | agggcatacac | ccccagcgtg | gttacataca | ccacactgat  | cgacggactg | 1200 |
| tgtaaagccg | gcgacgtgga  | cgaagccctc | gagctgttca | aagagatgcg  | gagcaagggc | 1260 |
| gtgaagccca | acgtgggtcac | atacaccacc | ctgatcgacg | gcctgtgcaa  | ggccggcaag | 1320 |
| gtggatgagg | ccctggagct  | gttcgacgag | atgaaggaga | ggggcatcaa  | gcccgcgctg | 1380 |
| gttacataca | acaccctgat  | cgacggcctg | tgcaagagcg | gcaagatcga  | ggaggccctg | 1440 |
| aagctgttca | aggagatgga  | ggagaagggc | atcaccccca | gcgtgggtcac | atacaccaca | 1500 |
| ctgatcgacg | gactgtgtaa  | agccggcgac | gtggacgaag | ccctcgagct  | gttcaaagag | 1560 |
| atgcggagca | agggcgtgaa  | gccccacgtg | gtgacataca | ccaccctgat  | cgacggcctg | 1620 |
| tgcaaggccg | gcaaagtgga  | cgaggccctg | gagctgttcg | acgagatgaa  | ggagaggggc | 1680 |
| atcaagcccc | acgtgggtcac | atacaacacc | ctgatcgacg | gcctgtgcaa  | gagcggcaag | 1740 |
| atcgaggagg | ccctgaagct  | gttcaaggag | atggaggaga | agggcatacac | ccccagcgtg | 1800 |
| gttacataca | ccacactgat  | cgacggactg | tgtaaagccg | gcgacgtgga  | cgaagccctc | 1860 |
| gagctgttca | aagagatgcg  | gagcaagggc | gtgaagccca | acgtgggtcac | atacaccacc | 1920 |
| ctgatcgacg | gcctgtgcaa  | ggccggcaag | gtggatgagg | ccctggagct  | gttcgacgag | 1980 |
| atgaaggaga | ggggcatcaa  | gcccgcgctg | gtcacataca | acaccctgat  | cgacggcctg | 2040 |
| tgcaagagcg | gcaagatcga  | ggaggccctg | aagctgttca | aggagatgga  | ggagaagggc | 2100 |
| atcaccccca | gcgtgggtcac | atacaccaca | ctgatcgacg | gactgtgtaa  | agccggcgac | 2160 |
| gtggacgaag | ccctcgagct  | gttcaaagag | atgcggagca | agggcgtgaa  | gccccacgtg | 2220 |
| gttacataca | ccaccctgat  | cgacggcctg | tgcaaggccg | gcaaagtgga  | cgaggccctg | 2280 |
| gagctgttcg | acgagatgaa  | ggagaggggc | atcaagcccc | acgtgggtcac | atacaacacc | 2340 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| ctgatcgacg gcctgtgcaa gagcggcaag atcgaggagg ccctgaagct gttcaaggag   | 2400 |
| atggaggaga agggcatcac ccccgacgtg gtgacataca ccacactgat cgacggactg   | 2460 |
| tgtaaagccg gcgacgtgga cgaagccctc gagctgttca aagagatgcg gagcaagggc   | 2520 |
| gtgaagccca acgtgggtcac atacaccacc ctgatcgacg gcctgtgcaa ggccggcaag  | 2580 |
| gtggatgagg ccctggagct gttcgacgag atgaaggaga ggggcatcaa gcccgcagag   | 2640 |
| ctgacctaca acaccctgat cagcggcctg ggcaaggccg gcagagccag agaccccccc   | 2700 |
| gtgctcagta gcccgaagaa gaaacgcaaa gtcgaggatc caaagaagaa aaggaaggtt   | 2760 |
| gaagacccca agaaaaagag gaaggtgggt tccgactata aggaccacga cggagactac   | 2820 |
| aaggatcatg atattgatta caaagacgat gacgataaga tggccccaaa gaagaagcgg   | 2880 |
| aaggtcggac gctagmagvs kgtgvvvdgd vnghksvrgg gdatngktkc ttgkvwtvtt   | 2940 |
| gygvacsryd hmkhdksamg yvrtskddgt yktravkgdt vnrkgdkdgn ghkynnshyv   | 3000 |
| ytadkkncka nkrhnvdgsv adhyntgdgv dnhyshsksk dnkrdhmvvt aagthgmdyk   | 3060 |
| srhmgnsvvt yntddgcksg kakkmkgtsv vtytt dgcka gdvdakmrsk gvknvvttytt | 3120 |
| dgckagkvda dmkrkgdvvt yntdgcksgk akkmkgtsvv tytt dgckag dvdakmrskg  | 3180 |
| vknvvttytt dgckagkvdad mkrkgdvvty ntdgcksgka kkmkgtsvvt ytt dgckagd | 3240 |
| vdakmrskgv knvvttytt dgckagkvdadm krgkdvvtynt dgcksgkak kmkgtsvvty  | 3300 |
| tt dgckagdv dakmrskgvk nvvttytt dgckagkvdadmk rgkdvvtynt dgcksgkakk | 3360 |
| mkgtsvvtyt dgckagdvda akmrskgvkn vvttytt dgckagkvdadmk rgkdvvtyntd  | 3420 |
| gcksgkakkm kgtsvvtytt dgckagdvda kmrskgvknv vtytt dgcka gkvdadmkrg  | 3480 |
| kdyntsggk agrardvssk kkrkvdkkkv kvdkkkkrkv sdykdhdgdy kdhd dykddd   | 3540 |
| dkmakkkrv gr                                                        | 3552 |

<210> 29  
 <211> 3189  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> NL\_PPRcag\_1

<220>  
 <221> misc\_feature  
 <222> (2626)..(2626)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2634)..(2634)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2644)..(2644)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2678)..(2678)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2698)..(2698)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2700)..(2700)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2704)..(2704)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2711)..(2711)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2740)..(2740)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2746)..(2746)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2787)..(2787)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2815)..(2815)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2856)..(2856)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2884)..(2884)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2925)..(2925)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2953)..(2953)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2994)..(2994)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3022)..(3022)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3063)..(3063)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3091)..(3091)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3132)..(3132)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (3158)..(3158)  
 <223> n is a, c, g, t or u

<400> 29  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac 60  
 aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggcgtg 120  
 agcgtcacc ccattccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc 180  
 cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc 240  
 ttcaaggtgg tgtatcccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc 300  
 ctctgtgatc acggcgtgac ccctaacatg atcgactact tcggcaggcc ctacgaggga 360  
 atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaaacaag 420  
 atcatcgacg agaggctgat caaccccgac ggctccctgc tgttcagggt gaccatcaat 480  
 ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg 540  
 cccagaggat ccggcggagg cggctccagc cggctccata tgggaaactc cgtgggtcaca 600  
 tacaacaccc tgatcgacgg cctgtgcaag agcggcaaga tcgaggaggc cctgaagctg 660  
 ttcaaggaga tggaggagaa gggcatcacc cccagcgtgg tcacatacac cacactgatc 720  
 gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg 780  
 agcaagggcg tgaagcccaa cgtggtgaca tacaccacc tgatcgacgg cctgtgcaag 840  
 gccggcaaag tggacgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag 900  
 cccgacgtgg tcacatacaa caccctgatc gacggcctgt gcaagagcgg caagatcgag 960  
 gaggccctga agctgttcaa ggagatggag gagaaggga tcacccccag cgtgggttaca 1020  
 tacaccacac tgatcgacgg actgtgtaaa gccggcgacg tggacgaagc cctcgagctg 1080  
 ttcaaagaga tgcggagcaa gggcgtgaag cccaacgtgg tcacatacac caccctgatc 1140  
 gacggcctgt gcaaggccgg caaggtggat gaggccctgg agctgttcga cgagatgaag 1200  
 gagaggggca tcaagcccga cgtgggttaca tacaacaccc tgatcgacgg cctgtgcaag 1260  
 agcggcaaga tcgaggaggc cctgaagctg ttcaaggaga tggaggagaa gggcatcacc 1320  
 cccagcgtgg tcacatacac cacactgatc gacggactgt gtaaagccgg cgacgtggac 1380  
 gaagccctcg agctgttcaa agagatgcgg agcaagggcg tgaagcccaa cgtggtgaca 1440

|             |             |             |             |             |              |      |
|-------------|-------------|-------------|-------------|-------------|--------------|------|
| tacaccaccc  | tgatcgacgg  | cctgtgcaag  | gccggcaaag  | tggaacgaggc | cctggagctg   | 1500 |
| ttcgacgaga  | tgaaggagag  | gggcatcaag  | cccgacgtgg  | tcacatacaa  | caccctgac    | 1560 |
| gacggcctgt  | gcaagagcgg  | caagatcgag  | gaggccctga  | agctgttcaa  | ggagatggag   | 1620 |
| gagaagggca  | tcacccccag  | cgtggttaca  | tacaccacac  | tgatcgacgg  | actgtgtaaa   | 1680 |
| gccggcgacg  | tggaacgaagc | cctcgagctg  | ttcaaagaga  | tgcggagcaa  | gggcgtgaag   | 1740 |
| cccaacgtgg  | tcacatacac  | caccctgac   | gacggcctgt  | gcaaggccgg  | caaggtggat   | 1800 |
| gaggccctgg  | agctgttcga  | cgagatgaag  | gagaggggca  | tcaagcccga  | cgtggtcaca   | 1860 |
| tacaacaccc  | tgatcgacgg  | cctgtgcaag  | agcggcaaga  | tcgaggaggc  | cctgaagctg   | 1920 |
| ttcaaggaga  | tggaaggagaa | gggcatcacc  | cccagcgtgg  | tcacatacac  | cacactgac    | 1980 |
| gacggactgt  | gtaaagccgg  | cgacgtggac  | gaagccctcg  | agctgttcaa  | agagatgcgg   | 2040 |
| agcaagggcg  | tgaagcccaa  | cgtggttaca  | tacaccaccc  | tgatcgacgg  | cctgtgcaag   | 2100 |
| gccggcaaag  | tggaacgaggc | cctggagctg  | ttcgacgaga  | tgaaggagag  | gggcatcaag   | 2160 |
| cccgacgtgg  | tcacatacaa  | caccctgac   | gacggcctgt  | gcaagagcgg  | caagatcgag   | 2220 |
| gaggccctga  | agctgttcaa  | ggagatggag  | gagaagggca  | tcacccccag  | cgtggtgaca   | 2280 |
| tacaccacac  | tgatcgacgg  | actgtgtaaa  | gccggcgacg  | tggaacgaagc | cctcgagctg   | 2340 |
| ttcaaagaga  | tgcggagcaa  | gggcgtgaag  | cccaacgtgg  | tcacatacac  | caccctgac    | 2400 |
| gacggcctgt  | gcaaggccgg  | caaggtggat  | gaggccctgg  | agctgttcga  | cgagatgaag   | 2460 |
| gagaggggca  | tcaagcccga  | cgagctgacc  | tacaacaccc  | tgatcagcgg  | cctgggcaag   | 2520 |
| gccggcagag  | ccagagaccc  | ccccgtgctc  | agtagcggct  | ccggcggcag  | cgggggaggc   | 2580 |
| gggcatcatc  | accatcacca  | cggacgctag  | magvtdvgdw  | rtagyndvvg  | vssngvsvtr   | 2640 |
| vsgngkdhvy  | gsgdmgkkvv  | yvddhhkvhy  | gtvdgvtnmd  | ygrygavdgk  | ktvtgtwngn   | 2700 |
| kdrndgsrvt  | ngvtgwrcra  | ggggsvrgsg  | gggssrhmg   | svvtyntdgc  | ksgkakkmkg   | 2760 |
| tsvvtytt dg | ckagdvdakm  | rskgvknvvt  | ytt dgckagk | vdadmkr gkd | vvtyntdgck   | 2820 |
| sgkakkmkg   | svvtytt dg  | kagdvdakmr  | skgvknvvt   | tt dgckagkv | dadmkr gkdv  | 2880 |
| vyntdgcks   | gkakkmkgts  | vvtytt dgck | agdvdakmrs  | kgvknvvt    | tyt dgckagkv | 2940 |
| admkr gkdvv | tyntdgcksg  | kakkmkgtsv  | vytt dgcka  | gdvdakmrsk  | gvknvvt      | 3000 |
| dgckagkvda  | dmkr gkdvv  | tyntdgcksgk | akkmkgtsv   | tytt dgckag | dvdakmrskg   | 3060 |

vknvtyttdd gckagkvdad mkrkgdvvty ntdgcksgka kkmkgtsvvt yttgdckagd 3120  
vdakmrskgv knvtyttddg ckagkvdadm krgkdynts ggkagrardv ssgsggsggg 3180  
ghhhhhhgr 3189

<210> 30  
<211> 3188  
<212> DNA  
<213> Artificial sequence

<220>  
<223> NL\_PPRcag\_2

<220>  
<221> misc\_feature  
<222> (2626)..(2626)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2634)..(2634)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2644)..(2644)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2678)..(2678)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2698)..(2698)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2700)..(2700)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2704)..(2704)  
<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (2711)..(2711)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2740)..(2740)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2746)..(2746)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2786)..(2786)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2814)..(2814)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2855)..(2855)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2883)..(2883)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2924)..(2924)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2952)..(2952)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2993)..(2993)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3021)..(3021)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3062)..(3062)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3090)..(3090)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3131)..(3131)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3157)..(3157)  
 <223> n is a, c, g, t or u

<400> 30  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac 60  
 aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggctgtg 120  
 agcgtcacc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc 180  
 cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc 240  
 ttcaaggtgg tgtatcccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc 300  
 ctcgtgatcg acggcgtgac ccctaacatg atcgactact tcggcaggcc ctacgaggga 360  
 atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaacaag 420  
 atcatcgacg agaggctgat caaccccgac ggctccctgc tgttcagggt gaccatcaat 480  
 ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg 540  
 cccagaggat ccggcggagg cggctccagc cggctccata tgggaaactc cgtggtcaca 600  
 tacaacacc tgatcgacga actgtgcaag agcggcaaga tcgaggaggc cctgaagctg 660  
 ttcaaggaga tggaggagaa gggcatcacc cccagcgtgg tcacatacac cacactgatc 720  
 gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg 780  
 agcaagggcg tgaagcccaa cgtggtgaca tacaccacc tgatcgacgg cctgtgcaag 840  
 gccggcaaag tggacgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag 900  
 cccgacgtgg tcacatacaa caccctgatc gacggcctgt gcaagagcgg caagatcgag 960

|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| gaggccctga | agctgttcaa  | ggagatggag  | gagaagggca | tcaccccag   | cgtgggttaca | 1020 |
| tacaccacac | tgatcgacgg  | actgtgtaaa  | gccggcgacg | tggaacgaagc | cctcgagctg  | 1080 |
| ttcaaagaga | tgaggagcaa  | gggcgtgaag  | cccaacgtgg | tcacatacac  | caccctgatac | 1140 |
| gacggcctgt | gcaaggccgg  | caaggtggat  | gaggccctgg | agctgttcga  | cgagatgaag  | 1200 |
| gagaggggca | tcaagcccga  | cgtgggttaca | tacaacaccc | tgatcgacgg  | cctgtgcaag  | 1260 |
| agcggcaaga | tcgaggaggc  | cctgaagctg  | ttcaaggaga | tgaggagagaa | gggcatcacc  | 1320 |
| cccagcgtgg | tcacatacac  | cacactgatac | gacggactgt | gtaaagccgg  | cgacgtggac  | 1380 |
| gaagccctcg | agctgttcaa  | agagatgcgg  | agcaagggcg | tgaagcccaa  | cgtgggtgaca | 1440 |
| tacaccaccc | tgatcgacgg  | cctgtgcaag  | gccggcaaag | tggaacgaggc | cctggagctg  | 1500 |
| ttcgacgaga | tgaaggagag  | gggcatcaag  | cccgacgtgg | tcacatacaa  | caccctgatac | 1560 |
| gacggcctgt | gcaagagcgg  | caagatcgag  | gaggccctga | agctgttcaa  | ggagatggag  | 1620 |
| gagaagggca | tcaccccag   | cgtgggttaca | tacaccacac | tgatcgacgg  | actgtgtaaa  | 1680 |
| gccggcgacg | tggaacgaagc | cctcgagctg  | ttcaaagaga | tgaggagcaa  | gggcgtgaag  | 1740 |
| cccaacgtgg | tcacatacac  | caccctgatac | gacggcctgt | gcaaggccgg  | caaggtggat  | 1800 |
| gaggccctgg | agctgttcga  | cgagatgaag  | gagaggggca | tcaagcccga  | cgtgggtcaca | 1860 |
| tacaacaccc | tgatcgacgg  | cctgtgcaag  | agcggcaaga | tcgaggaggc  | cctgaagctg  | 1920 |
| ttcaaggaga | tgaggagagaa | gggcatcacc  | cccagcgtgg | tcacatacac  | cacactgatac | 1980 |
| gacggactgt | gtaaagccgg  | cgacgtggac  | gaagccctcg | agctgttcaa  | agagatgcgg  | 2040 |
| agcaagggcg | tgaagcccaa  | cgtgggttaca | tacaccaccc | tgatcgacgg  | cctgtgcaag  | 2100 |
| gccggcaaag | tggaacgaggc | cctggagctg  | ttcgacgaga | tgaaggagag  | gggcatcaag  | 2160 |
| cccgacgtgg | tcacatacaa  | caccctgatac | gacggcctgt | gcaagagcgg  | caagatcgag  | 2220 |
| gaggccctga | agctgttcaa  | ggagatggag  | gagaagggca | tcaccccag   | cgtgggtgaca | 2280 |
| tacaccacac | tgatcgacgg  | actgtgtaaa  | gccggcgacg | tggaacgaagc | cctcgagctg  | 2340 |
| ttcaaagaga | tgaggagcaa  | gggcgtgaag  | cccaacgtgg | tcacatacac  | caccctgatac | 2400 |
| gacggcctgt | gcaaggccgg  | caaggtggat  | gaggccctgg | agctgttcga  | cgagatgaag  | 2460 |
| gagaggggca | tcaagcccga  | cgagctgacc  | tacaacaccc | tgatcagcgg  | cctgggcaag  | 2520 |

```

gccggcagag ccagagaccc ccccgctgctc agtagcggct ccggcggcag cgggggaggc 2580
gggcatcatc accatcacca cggacgctag magvtdvgdw rtagyndvgg vssngvsvtr 2640
vsgngkdhvy gsgdmgkkvv yvddhhkvhy gtvdgvtnmd ygrygavdgk ktvtgtwngn 2700
kdrndgsrvt ngvtgwrcra ggggsvrgsg gggssrhmgm svvtyntdck sgkakkmkgt 2760
svvtyttddgc kagdvdakmr skgvknvvtty ttgdgckagkv dadmkrkdv vtyntdgcks 2820
gkakkmkgtts vvtyttddgck agdvdakmrs kgvknvvttyt tdgckagkv admkrkdvv 2880
tyntdgcksg kakkmkgtsv vtyttddgcka gdvdakmrsk gvknvvttyt dgckagkvda 2940
dmkrkdvvty yntdgcksgk akkmkgtsvv tyttddgckag dvdakmrskg vknvvttytt 3000
gckagkvdad mkrkdvvty ntdgcksgka kkmkgtsvvt yttddgckagd vdakmrskgv 3060
knvvttyttg ckagkvdadm krgkdvvtyt ntdgcksgkak kmkgtsvvtty ttddgckagdv 3120
dakmrskgvk nvvtyttddgc kagkvdadmk rgkdyntsg gkagrardvs sgsggsgggg 3180
hhhhhghr 3188

```

```

<210> 31
<211> 3189
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> NL_PPRcag_3

```

```

<220>
<221> misc_feature
<222> (2626)..(2626)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2634)..(2634)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2644)..(2644)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2678)..(2678)
<223> n is a, c, g, t or u

```

<220>  
 <221> misc\_feature  
 <222> (2698)..(2698)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2700)..(2700)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2704)..(2704)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2711)..(2711)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2740)..(2740)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2746)..(2746)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2748)..(2748)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2787)..(2787)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2815)..(2815)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2856)..(2856)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (2884)..(2884)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2953)..(2953)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2994)..(2994)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3022)..(3022)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3063)..(3063)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3091)..(3091)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3132)..(3132)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3158)..(3158)  
 <223> n is a, c, g, t or u

<400> 31  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac 60  
 aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggctg 120  
 agcgtcacc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc 180  
 cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc 240  
 ttcaaggtgg tgtatccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc 300

|             |            |             |            |            |             |      |
|-------------|------------|-------------|------------|------------|-------------|------|
| ctcgtgatcg  | acggcgtgac | ccctaacatg  | atcgactact | tcggcaggcc | ctacgaggga  | 360  |
| atcgccgtgt  | tcgacggaaa | gaagatcacc  | gtgaccggca | ccctgtggaa | cggaaacaag  | 420  |
| atcatcgacg  | agaggctgat | caaccccgac  | ggctccctgc | tgttcagggt | gaccatcaat  | 480  |
| ggcgtgaccg  | gctggagact | gtgcgagaga  | atcctggccg | gaggcggagg | aagcctcgtg  | 540  |
| cccagaggat  | ccggcggagg | cggctccagc  | cggctccata | tgggaaactc | cgtggtcaca  | 600  |
| tacaacacca  | acatcgacca | gctgtgcaag  | agcggcaaga | tcgaggaggc | cctgaagctg  | 660  |
| ttcaaggaga  | tggaggagaa | gggcatcacc  | cccagcgtgg | tcacatacac | cacactgatac | 720  |
| gacggactgt  | gtaaagccgg | cgacgtggac  | gaagccctcg | agctgttcaa | agagatgcgg  | 780  |
| agcaagggcg  | tgaagcccaa | cgtggtgaca  | tacaccaccc | tgatcgacgg | cctgtgcaag  | 840  |
| gccggcaaag  | tggacgaggc | cctggagctg  | ttcgacgaga | tgaaggagag | gggcatcaag  | 900  |
| cccgacgtgg  | tcacatacaa | caccctgatac | gacggcctgt | gcaagagcgg | caagatcgag  | 960  |
| gaggccctga  | agctgttcaa | ggagatggag  | gagaagggca | tcacccccag | cgtggttaca  | 1020 |
| tacaccacac  | tgatcgacgg | actgtgtaaa  | gccggcgacg | tggacgaagc | cctcgagctg  | 1080 |
| ttcaaagaga  | tgcggagcaa | gggcgtgaag  | cccaacgtgg | tcacatacac | caccctgatac | 1140 |
| gacggcctgt  | gcaaggccgg | caaggtggat  | gaggccctgg | agctgttcga | cgagatgaag  | 1200 |
| gagagggggca | tcaagcccga | cgtggttaca  | tacaacaccc | tgatcgacgg | cctgtgcaag  | 1260 |
| agcggcaaga  | tcgaggaggc | cctgaagctg  | ttcaaggaga | tggaggagaa | gggcatcacc  | 1320 |
| cccagcgtgg  | tcacatacac | cacactgatac | gacggactgt | gtaaagccgg | cgacgtggac  | 1380 |
| gaagccctcg  | agctgttcaa | agagatgcgg  | agcaagggcg | tgaagcccaa | cgtggtgaca  | 1440 |
| tacaccaccc  | tgatcgacgg | cctgtgcaag  | gccggcaaag | tggacgaggc | cctggagctg  | 1500 |
| ttcgacgaga  | tgaaggagag | gggcatcaag  | cccgacgtgg | tcacatacaa | caccctgatac | 1560 |
| gacggcctgt  | gcaagagcgg | caagatcgag  | gaggccctga | agctgttcaa | ggagatggag  | 1620 |
| gagaagggca  | tcacccccag | cgtggttaca  | tacaccacac | tgatcgacgg | actgtgtaaa  | 1680 |
| gccggcgacg  | tggacgaagc | cctcgagctg  | ttcaaagaga | tgcggagcaa | gggcgtgaag  | 1740 |
| cccaacgtgg  | tcacatacac | caccctgatac | gacggcctgt | gcaaggccgg | caaggtggat  | 1800 |
| gaggccctgg  | agctgttcga | cgagatgaag  | gagaggggca | tcaagcccga | cgtggtcaca  | 1860 |

|             |            |              |             |              |              |      |
|-------------|------------|--------------|-------------|--------------|--------------|------|
| tacaacaccc  | tgatcgacgg | cctgtgcaag   | agcggcaaga  | tcgaggaggc   | cctgaagctg   | 1920 |
| ttcaaggaga  | tggaggagaa | gggcatcacc   | cccagcgtgg  | tcacatacac   | cacactgac    | 1980 |
| gacggactgt  | gtaaagccgg | cgacgtggac   | gaagccctcg  | agctgttcaa   | agagatgcgg   | 2040 |
| agcaagggcg  | tgaagcccaa | cgtgggttaca  | tacaccaccc  | tgatcgacgg   | cctgtgcaag   | 2100 |
| gccggcaaag  | tggacgaggc | cctggagctg   | ttcgacgaga  | tgaaggagag   | gggcatcaag   | 2160 |
| cccagcgtgg  | tcacatacaa | caccctgac    | gacggcctgt  | gcaagagcgg   | caagatcgag   | 2220 |
| gaggccctga  | agctgttcaa | ggagatggag   | gagaagggca  | tcacccccag   | cgtggtgaca   | 2280 |
| tacaccacac  | tgatcgacgg | actgtgtaaa   | gccggcgacg  | tggacgaagc   | cctcgagctg   | 2340 |
| ttcaaagaga  | tgcggagcaa | gggcgtgaag   | cccaacgtgg  | tcacatacac   | caccctgac    | 2400 |
| gacggcctgt  | gcaaggccgg | caaggtggat   | gaggccctgg  | agctgttcga   | cgagatgaag   | 2460 |
| gagaggggca  | tcaagcccga | cgagctgacc   | tacaacaccc  | tgatcagcgg   | cctgggcaag   | 2520 |
| gccggcagag  | ccagagaccc | ccccgtgctc   | agtagcggct  | ccggcggcag   | cgggggaggc   | 2580 |
| gggcatcatc  | accatcacca | cggacgctag   | magvtdvgdw  | rtagyndvgg   | vssngvsvtr   | 2640 |
| vsgngkdhvy  | gsgdmgkkvv | yvddhhkvhy   | gtvdgvtnmd  | ygrygavdgk   | ktvtgtwngn   | 2700 |
| kdrndgsrvt  | ngvtgwrcra | ggggsvrgsg   | gggssrhmg   | svvtyntndc   | ksgkakkmkg   | 2760 |
| tsvvtytt dg | ckagdvdakm | rskgvknvvt   | ytt dgckagk | vdadmkr gkd  | vvtyntdgck   | 2820 |
| sgkakkmkg   | svvtytt dg | kagdvdakmr   | skgvknvvt   | tytt dgckagk | vdadmkr gkd  | 2880 |
| tyntdgcks   | gkakkmkgts | vvtytt dgck  | agdvdakmrs  | kgvknvvt     | tytt dgckagk | 2940 |
| admkr gkd   | tyntdgcksg | kakkmkgts    | vtytt dgcka | gdvdakmrsk   | gvknvvt      | 3000 |
| dgckagkvda  | dmkr gkd   | tyntdgcksgk  | akkmkgts    | tytt dgckag  | gdvdakmrskg  | 3060 |
| vkvnvtytt   | dgckagkvda | dmkr gkd     | tyntdgcksgk | akkmkgts     | tytt dgckag  | 3120 |
| vdakmrskg   | knvvt      | tytt dgckagk | vdadmkr gkd | tyntdgcksgk  | akkmkgts     | 3180 |
| ghhhhhgr    |            |              |             |              |              | 3189 |

&lt;210&gt; 32

&lt;211&gt; 3189

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

<223> NL\_PPRcag\_4

<220>

<221> misc\_feature

<222> (2626)..(2626)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2634)..(2634)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2644)..(2644)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2678)..(2678)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2698)..(2698)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2700)..(2700)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2704)..(2704)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2711)..(2711)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2740)..(2740)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2746)..(2746)

<223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2748)..(2748)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2787)..(2787)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2815)..(2815)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2856)..(2856)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2884)..(2884)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2925)..(2925)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2953)..(2953)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2994)..(2994)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3022)..(3022)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3063)..(3063)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3091)..(3091)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3132)..(3132)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (3158)..(3158)

<223> n is a, c, g, t or u

<400> 32

|                                                                    |      |
|--------------------------------------------------------------------|------|
| atggccggag tggtcacact ggaggacttc gtgggcgact ggagacagac cgccggctac  | 60   |
| aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggctgtg | 120  |
| agcgtcacc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc   | 180  |
| cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc  | 240  |
| ttcaaggtgg tgtatccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc   | 300  |
| ctcgtgatcg acggcgtgac ccctaactg atcgactact tcggcaggcc ctacgaggga   | 360  |
| atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaaacaag  | 420  |
| atcatcgacg agaggctgat caaccccgac ggctccctgc tggtcagggt gaccatcaat  | 480  |
| ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg  | 540  |
| cccagaggat ccggcggagg cggctccagc cggctccata tgggaaactc cgtggtcaca  | 600  |
| tacaacacca acatcgacga actgtgcaag agcggcaaga tcgaggaggc cctgaagctg  | 660  |
| ttcaaggaga tggaggagaa gggcatcacc cccagcgtgg tcacatacac cacactgac   | 720  |
| gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg  | 780  |
| agcaagggcg tgaagcccaa cgtggtgaca tacaccacc tgatcgacgg cctgtgcaag   | 840  |
| gccggcaaag tggacgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag  | 900  |
| cccgacgtgg tcacatacaa caccctgac gacggcctgt gcaagagcgg caagatcgag   | 960  |
| gaggccctga agctgttcaa ggagatggag gagaagggca tcacccccag cgtggttaca  | 1020 |
| tacaccacac tgatcgacgg actgtgtaaa gccggcgacg tggacgaagc cctcgagctg  | 1080 |
| ttcaaagaga tgcggagcaa gggcgtgaag cccaacgtgg tcacatacac caccctgac   | 1140 |
| gacggcctgt gcaaggccgg caaggtggat gaggccctgg agctgttcga cgagatgaag  | 1200 |

|             |            |             |             |             |             |      |
|-------------|------------|-------------|-------------|-------------|-------------|------|
| gagaggggca  | tcaagcccga | cgtgggttaca | tacaacaccc  | tgatcgacgg  | cctgtgcaag  | 1260 |
| agcggcaaga  | tcgaggaggc | cctgaagctg  | ttcaaggaga  | tggaggagaa  | gggcatcacc  | 1320 |
| cccagcgtgg  | tcacatacac | cacactgatac | gacggactgt  | gtaaagccgg  | cgacgtggac  | 1380 |
| gaagccctcg  | agctgttcaa | agagatgcgg  | agcaagggcg  | tgaagcccaa  | cgtggtgaca  | 1440 |
| tacaccaccc  | tgatcgacgg | cctgtgcaag  | gccggcaaag  | tggacgaggc  | cctggagctg  | 1500 |
| ttcgacgaga  | tgaaggagag | gggcatcaag  | cccgacgtgg  | tcacatacaa  | caccctgatac | 1560 |
| gacggcctgt  | gcaagagcgg | caagatcgag  | gaggccctga  | agctgttcaa  | ggagatggag  | 1620 |
| gagaagggca  | tcacccccag | cgtgggttaca | tacaccacac  | tgatcgacgg  | actgtgtaaa  | 1680 |
| gccggcgacg  | tggacgaagc | cctcgagctg  | ttcaaagaga  | tgcggagcaa  | gggcgtgaag  | 1740 |
| cccaacgtgg  | tcacatacac | caccctgatac | gacggcctgt  | gcaaggccgg  | caaggtggat  | 1800 |
| gaggccctgg  | agctgttcga | cgagatgaag  | gagaggggca  | tcaagcccga  | cgtggtcaca  | 1860 |
| tacaacaccc  | tgatcgacgg | cctgtgcaag  | agcggcaaga  | tcgaggaggc  | cctgaagctg  | 1920 |
| ttcaaggaga  | tggaggagaa | gggcatcacc  | cccagcgtgg  | tcacatacac  | cacactgatac | 1980 |
| gacggactgt  | gtaaagccgg | cgacgtggac  | gaagccctcg  | agctgttcaa  | agagatgcgg  | 2040 |
| agcaagggcg  | tgaagcccaa | cgtgggttaca | tacaccaccc  | tgatcgacgg  | cctgtgcaag  | 2100 |
| gccggcaaag  | tggacgaggc | cctggagctg  | ttcgacgaga  | tgaaggagag  | gggcatcaag  | 2160 |
| cccgacgtgg  | tcacatacaa | caccctgatac | gacggcctgt  | gcaagagcgg  | caagatcgag  | 2220 |
| gaggccctga  | agctgttcaa | ggagatggag  | gagaagggca  | tcacccccag  | cgtggtgaca  | 2280 |
| tacaccacac  | tgatcgacgg | actgtgtaaa  | gccggcgacg  | tggacgaagc  | cctcgagctg  | 2340 |
| ttcaaagaga  | tgcggagcaa | gggcgtgaag  | cccaacgtgg  | tcacatacac  | caccctgatac | 2400 |
| gacggcctgt  | gcaaggccgg | caaggtggat  | gaggccctgg  | agctgttcga  | cgagatgaag  | 2460 |
| gagaggggca  | tcaagcccga | cgagctgacc  | tacaacaccc  | tgatcagcgg  | cctgggcaag  | 2520 |
| gccggcagag  | ccagagaccc | ccccgtgctc  | agtagcggct  | ccggcggcag  | cgggggaggc  | 2580 |
| gggcatcatc  | accatcacca | cggacgctag  | magvtdvgdw  | rtagyndvvg  | vssngvsvtr  | 2640 |
| vsgngkdhvy  | gsgdmgkkvv | yvddhhkvhy  | gtvdgvtnmd  | ygrygavdgk  | ktvtgtwngn  | 2700 |
| kdrndgsrvt  | ngvtgwrcra | ggggsvrgsg  | gggssrhmgm  | svvtyntndc  | ksgkakkmkg  | 2760 |
| tsvvtytt dg | ckagdvdakm | rskgvknvvt  | ytt dgckagk | vdadmkr gkd | vvtyntdgck  | 2820 |

```

sgkakkmkgt svvtyttgdc kagdvda kmr skgvknvvt y ttdgckagkv dadmkr gkdv 2880
vtyntdgcks gkakkmkgt s vvyttgdgck agdvda kmr s kgvknvvt y t dgckagkv d 2940
admkr gkdvv tyntdgck s g kakkmkgt s v vtyttgdgck a gdvda kmr s k g vknvvt y t t 3000
dgckagkv da dmkr gkdvv t yntdgck s g k akkmkg t s v v tyttgdgck a g dvda kmr s k g 3060
vknvvt y t t d gckagkv da d mkr gkdvv t y ntdgck s g k kkmkg t s v v y t t d gckag d 3120
vdakmr s kgv knvvt y t t d g ckagkv da d m krgk d t y n t s g g k a g r a r d v s s g s g g s g g g 3180
ghhhhhhgr 3189

```

```

<210> 33
<211> 3190
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> NL_PPRcag_5

```

```

<220>
<221> misc_feature
<222> (2626)..(2626)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2634)..(2634)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2644)..(2644)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2678)..(2678)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2698)..(2698)
<223> n is a, c, g, t or u

```

```

<220>
<221> misc_feature
<222> (2700)..(2700)

```

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2704)..(2704)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2711)..(2711)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2740)..(2740)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2746)..(2746)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2748)..(2748)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2788)..(2788)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2816)..(2816)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2857)..(2857)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2885)..(2885)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature

<222> (2926)..(2926)

<223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (2954)..(2954)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2995)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3064)..(3064)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3092)..(3092)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3133)..(3133)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3159)..(3159)  
 <223> n is a, c, g, t or u

<400> 33  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac 60  
 aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggcgtg 120  
 agcgtcacc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc 180  
 cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc 240  
 ttcaaggtgg tgtatcccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc 300  
 ctcgtgatcg acggcgtgac ccctaacaatg atcgactact tcggcaggcc ctacgaggga 360  
 atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaaacaag 420  
 atcatcgacg agaggctgat caaccccgac ggctccctgc tgttcagggt gaccatcaat 480  
 ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg 540

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| cccagaggat | ccggcggagg | cggctccagc | cggctccata | tgggaaactc | cgtggtcaca | 600  |
| tacaacacca | acatcgacaa | actgtgcaag | agcggcaaga | tcgaggaggc | cctgaagctg | 660  |
| ttcaaggaga | tggaggagaa | gggcatcacc | cccagcgtgg | tcacatacac | cacactgatc | 720  |
| gacggactgt | gtaaagccgg | cgacgtggac | gaagccctcg | agctgttcaa | agagatgcgg | 780  |
| agcaagggcg | tgaagcccaa | cgtggtgaca | tacaccaccc | tgatcgacgg | cctgtgcaag | 840  |
| gccggcaaag | tggacgaggc | cctggagctg | ttcgacgaga | tgaaggagag | gggcatcaag | 900  |
| cccgacgtgg | tcacatacaa | caccctgatc | gacggcctgt | gcaagagcgg | caagatcgag | 960  |
| gaggccctga | agctgttcaa | ggagatggag | gagaagggca | tcacccccag | cgtggttaca | 1020 |
| tacaccacac | tgatcgacgg | actgtgtaaa | gccggcgacg | tggacgaagc | cctcgagctg | 1080 |
| ttcaaagaga | tgcggagcaa | gggcgtgaag | cccaacgtgg | tcacatacac | caccctgatc | 1140 |
| gacggcctgt | gcaaggccgg | caaggtggat | gaggccctgg | agctgttcga | cgagatgaag | 1200 |
| gagaggggca | tcaagcccga | cgtggttaca | tacaacaccc | tgatcgacgg | cctgtgcaag | 1260 |
| agcggcaaga | tcgaggaggc | cctgaagctg | ttcaaggaga | tggaggagaa | gggcatcacc | 1320 |
| cccagcgtgg | tcacatacac | cacactgatc | gacggactgt | gtaaagccgg | cgacgtggac | 1380 |
| gaagccctcg | agctgttcaa | agagatgcgg | agcaagggcg | tgaagcccaa | cgtggtgaca | 1440 |
| tacaccaccc | tgatcgacgg | cctgtgcaag | gccggcaaag | tggacgaggc | cctggagctg | 1500 |
| ttcgacgaga | tgaaggagag | gggcatcaag | cccgacgtgg | tcacatacaa | caccctgatc | 1560 |
| gacggcctgt | gcaagagcgg | caagatcgag | gaggccctga | agctgttcaa | ggagatggag | 1620 |
| gagaagggca | tcacccccag | cgtggttaca | tacaccacac | tgatcgacgg | actgtgtaaa | 1680 |
| gccggcgacg | tggacgaagc | cctcgagctg | ttcaaagaga | tgcggagcaa | gggcgtgaag | 1740 |
| cccaacgtgg | tcacatacac | caccctgatc | gacggcctgt | gcaaggccgg | caaggtggat | 1800 |
| gaggccctgg | agctgttcga | cgagatgaag | gagaggggca | tcaagcccga | cgtggtcaca | 1860 |
| tacaacaccc | tgatcgacgg | cctgtgcaag | agcggcaaga | tcgaggaggc | cctgaagctg | 1920 |
| ttcaaggaga | tggaggagaa | gggcatcacc | cccagcgtgg | tcacatacac | cacactgatc | 1980 |
| gacggactgt | gtaaagccgg | cgacgtggac | gaagccctcg | agctgttcaa | agagatgcgg | 2040 |
| agcaagggcg | tgaagcccaa | cgtggttaca | tacaccaccc | tgatcgacgg | cctgtgcaag | 2100 |
| gccggcaaag | tggacgaggc | cctggagctg | ttcgacgaga | tgaaggagag | gggcatcaag | 2160 |

```

cccgacgtgg tcacatacaa caccctgacg gacggcctgt gcaagagcgg caagatcgag      2220
gaggccctga agctgttcaa ggagatggag gagaagggca tcacccccag cgtggtgaca      2280
tacaccacac tgatcgacgg actgtgtaaa gccggcgacg tggacgaagc cctcgagctg      2340
ttcaaagaga tgcggagcaa gggcgtgaag cccaacgtgg tcacatacac caccctgacg      2400
gacggcctgt gcaaggccgg caaggtggat gaggccctgg agctgttcga cgagatgaag      2460
gagaggggca tcaagcccga cgagctgacc tacaacaccc tgatcagcgg cctgggcaag      2520
gccggcagag ccagagaccc ccccgctgctc agtagcggct ccggcggcag cgggggaggc      2580
gggcatcatc accatcacca cggacgctag magvtdvgdw rtagyndvgg vssngvsvtr      2640
vsgngkdhvy gsgdmgkkvv yvddhhkvhy gtvdgvtnmd ygrygavdgk ktvtgtwngn      2700
kdrndgsrvt ngvtgwrcra ggggsvrgsg gggssrhmgm svvtyntndk ckskkakmk      2760
gtsvvtyttt gckagdvdak mrskgvknvv tytttgckag kvdadmkrk dvvtyntdgc      2820
kskkakmkkg tsvvtytttg ckagdvdakm rskgvknvvt ytttgckagk vdadmkrkdg      2880
vvtyntdgck sgkakmkgt svvtyttgdc kagdvdakmr skgvknvty ttdgckagkv      2940
dadmkrkdv vtyntdgcks gkakmkgt svvtyttgdc agdvdakmrs kgvknvtyt      3000
tdgckagkvd admkrkdvv tyntdgcksg kakkmkgtsv vtyttgcka gdvdakmrsk      3060
gvknvtytt dgckagkvda dmkrkdvvt yntdgcksgk akkmkgtsvv tyttgckag      3120
dvdakmrskg vknvtyttt gckagkvdad mkrkgttynt sggkagrard vsssgsgsgg      3180
gghhhhhhgr      3190

```

```

<210> 34
<211> 3190
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> NL_PPRcag_1

```

```

<220>
<221> misc_feature
<222> (2626)..(2626)
<223> n is a, c, g, t or u

```

```

<220>

```

<221> misc\_feature  
<222> (2634)..(2634)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2644)..(2644)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2678)..(2678)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2698)..(2698)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2700)..(2700)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2704)..(2704)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2711)..(2711)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2740)..(2740)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2746)..(2746)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2788)..(2788)  
<223> n is a, c, g, t or u

<220>  
<221> misc\_feature  
<222> (2816)..(2816)  
<223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2857)..(2857)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2885)..(2885)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2926)..(2926)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2954)..(2954)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (2995)..(2995)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3023)..(3023)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3064)..(3064)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3092)..(3092)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3133)..(3133)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (3159)..(3159)  
 <223> n is a, c, g, t or u

<400> 34  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac

|                                                                   |      |
|-------------------------------------------------------------------|------|
| aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggcgtg | 120  |
| agcgtcaccc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc | 180  |
| cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc | 240  |
| ttcaaggtgg tgtatcccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc | 300  |
| ctcgtgatcg acggcgtgac ccctaacatg atcgactact tcggcaggcc ctacgaggga | 360  |
| atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaaacaag | 420  |
| atcatcgacg agaggctgat caaccccgcg ggctccctgc tgttcagggt gaccatcaat | 480  |
| ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg | 540  |
| cccagaggat ccggcggagg cggtccagc cggctccata tgggaaactc cgtggtcaca  | 600  |
| tacaacaccg atatcgacgg cctgtgcaag agcggcaaga tcgaggaggc cctgaagctg | 660  |
| ttcaaggaga tggaggagaa gggcatcacc cccagcgtgg tcacatacac cacactgatc | 720  |
| gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg | 780  |
| agcaagggcg tgaagcccaa cgtggtgaca tacaccacc tgatcgacgg cctgtgcaag  | 840  |
| gccggcaaag tggacgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag | 900  |
| cccgacgtgg tcacatacaa caccctgatc gacggcctgt gcaagagcgg caagatcgag | 960  |
| gaggccctga agctgttcaa ggagatggag gagaagggca tcacccccag cgtggttaca | 1020 |
| tacaccacac tgatcgacgg actgtgtaaa gccggcgacg tggacgaagc cctcgagctg | 1080 |
| ttcaaagaga tgcggagcaa gggcgtgaag cccaacgtgg tcacatacac caccctgatc | 1140 |
| gacggcctgt gcaaggccgg caaggtggat gaggccctgg agctgttcga cgagatgaag | 1200 |
| gagaggggca tcaagcccga cgtgggttaca tacaacacc tgatcgacgg cctgtgcaag | 1260 |
| agcggcaaga tcgaggaggc cctgaagctg ttcaaggaga tggaggagaa gggcatcacc | 1320 |
| cccagcgtgg tcacatacac cacactgatc gacggactgt gtaaagccgg cgacgtggac | 1380 |
| gaagccctcg agctgttcaa agagatgcgg agcaagggcg tgaagcccaa cgtggtgaca | 1440 |
| tacaccaccc tgatcgacgg cctgtgcaag gccggcaaag tggacgaggc cctggagctg | 1500 |
| ttcgacgaga tgaaggagag gggcatcaag cccgacgtgg tcacatacaa caccctgatc | 1560 |
| gacggcctgt gcaagagcgg caagatcgag gaggccctga agctgttcaa ggagatggag | 1620 |

|            |             |             |             |             |             |      |
|------------|-------------|-------------|-------------|-------------|-------------|------|
| gagaagggca | tcacccccag  | cgtgggttaca | tacaccacac  | tgatcgacgg  | actgtgtaaa  | 1680 |
| gccggcgacg | tggaacgaagc | cctcgagctg  | ttcaaagaga  | tgcgagagcaa | gggcgtgaag  | 1740 |
| cccaacgtgg | tcacatacac  | caccctgatac | gacggcctgt  | gcaaggccgg  | caaggtggat  | 1800 |
| gaggccctgg | agctgttcga  | cgagatgaag  | gagaggggca  | tcaagcccga  | cgtggtcaca  | 1860 |
| tacaacaccc | tgatcgacgg  | cctgtgcaag  | agcggcaaga  | tcgaggaggc  | cctgaagctg  | 1920 |
| ttcaaggaga | tggaaggagaa | gggcatcacc  | cccagcgtgg  | tcacatacac  | cacactgatac | 1980 |
| gacggactgt | gtaaagccgg  | cgacgtggac  | gaagccctcg  | agctgttcaa  | agagatgcgg  | 2040 |
| agcaagggcg | tgaagcccaa  | cgtgggttaca | tacaccaccc  | tgatcgacgg  | cctgtgcaag  | 2100 |
| gccggcaaag | tggaacgaggc | cctggagctg  | ttcgacgaga  | tgaaggagag  | gggcatcaag  | 2160 |
| cccagcgtgg | tcacatacaa  | caccctgatac | gacggcctgt  | gcaagagcgg  | caagatcgag  | 2220 |
| gaggccctga | agctgttcaa  | ggagatggag  | gagaagggca  | tcacccccag  | cgtgggtgaca | 2280 |
| tacaccacac | tgatcgacgg  | actgtgtaaa  | gccggcgacg  | tggaacgaagc | cctcgagctg  | 2340 |
| ttcaaagaga | tgcgagagcaa | gggcgtgaag  | cccaacgtgg  | tcacatacac  | caccctgatac | 2400 |
| gacggcctgt | gcaaggccgg  | caaggtggat  | gaggccctgg  | agctgttcga  | cgagatgaag  | 2460 |
| gagaggggca | tcaagcccga  | cgagctgacc  | tacaacaccc  | tgatcagcgg  | cctgggcaag  | 2520 |
| gccggcagag | ccagagaccc  | ccccgtgctc  | agtagcggct  | ccggcggcag  | cgggggaggc  | 2580 |
| gggcatcatc | accatcacca  | cggacgctag  | magvtdvgdw  | rtagyndvvg  | vssngvsvtr  | 2640 |
| vsgngkdhvy | gsgdmgkkvv  | yvddhhkvhy  | gtvdgvtnmd  | ygrygavdgk  | ktvtgtwngn  | 2700 |
| kdrndgsrvt | ngvtgwrcra  | ggggsvrgsg  | gggssrhmg   | svvtyntddg  | cksgkakkmk  | 2760 |
| gtsvvtytt  | gckagdvda   | mrskgvknvv  | tytt dgckag | kvdadmkrk   | dvvtyntdgc  | 2820 |
| ksgkakkmkg | tsvvtytt dg | ckagdvda    | mrskgvknvv  | tytt dgckag | vdadmkrkd   | 2880 |
| vvtyntdgck | sgkakkmkg   | svvtytt dg  | kagdvda     | mrskgvknvv  | tytt dgckag | 2940 |
| admkrkd    | vvtyntdgck  | sgkakkmkg   | svvtytt dg  | agdvda      | mrskgvknvv  | 3000 |
| tdgckagkd  | admkrkd     | vvtyntdgck  | sgkakkmkg   | svvtytt dg  | agdvda      | 3060 |
| gvknvvtytt | dgckagkd    | admkrkd     | vvtyntdgck  | sgkakkmkg   | svvtytt dg  | 3120 |
| dvda       | mrskgvknvv  | tytt dgckag | vdadmkrkd   | tsvvtytt dg | ckagdvda    | 3180 |
| gghhhhhhgr |             |             |             |             |             | 3190 |

<210> 35  
 <211> 1066  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> GFPnls

<220>  
 <221> misc\_feature  
 <222> (878)..(878)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (934)..(934)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (946)..(946)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (954)..(954)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (959)..(959)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (961)..(961)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (964)..(964)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (972)..(972)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature

<222> (976)..(976)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (980)..(980)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (989)..(989)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (996)..(996)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1006)..(1006)  
 <223> n is a, c, g, t or u

<400> 35  
 atggccggag tgagcaaggg cgaggagctg ttcaccgggg tggtgcccac cctgggtcgag 60  
 ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc 120  
 acctacggca agctgaccct gaagttcatc tgcaccaccg gcaagctgcc cgtgccctgg 180  
 cccaccctcg tgaccaccct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac 240  
 atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc 300  
 atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac 360  
 accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg 420  
 gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag 480  
 aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg cagcgtgcag 540  
 ctgcccgacc actaccagca gaacaccccc atcggcgacg gccccgtgtc gctgcccgac 600  
 aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac 660  
 atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac 720  
 aagccaaaga aaaagagaaa ggtagccat ggctccggcg gcagcggggg aggtagcggg 780  
 gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac 840  
 gataagatgg ccggacgcta gmagvskgtg vvvdgdvngh ksvsgggdat ygktkcttgk 900

|                                                                   |      |
|-------------------------------------------------------------------|------|
| vwtvtttygv csrydhmkhd ksamgyvrtk ddgnyktrav kgdtvnrkgd kdgngkhyny | 960  |
| nshnvymadk kngkvnkrhn dgsvadhynt gdgvdnhyst saskdnkrdh mvvtaagtg  | 1020 |
| dykkkkkrkvs hgsggsgggs gdykdhdgdy kdhddykddd dkmagr               | 1066 |

<210> 36  
 <211> 1170  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> mClover3nls

<220>  
 <221> misc\_feature  
 <222> (959)..(959)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (972)..(972)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1029)..(1029)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1037)..(1037)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1042)..(1043)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1054)..(1054)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1058)..(1058)  
 <223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (1062)..(1062)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1072)..(1072)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1079)..(1079)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (1089)..(1089)  
 <223> n is a, c, g, t or u

<400> 36  
 atggccggag tgtccaaagg cgaggagctg tttaccggcg tcgtgcctat tctggtggag 60  
 ctggacggcg acgtgaacgg ccacaagttc tccgtgaggg gcgagggcga aggcgatgcc 120  
 acaaacggca agctgaccct caagttcatc tgcaccactg gtaaactgcc cgttccttgg 180  
 cccacactgg tgaccacctt cggctacggc gtggcttggt tctctcgta ccccgaccat 240  
 atgaagcagc acgacttctt caagtccgcc atgcccagg gatacgtgca agaaaggacc 300  
 atctccttca aggacgatgg cacctacaag accagagccg aggtgaagtt cgagggcgac 360  
 aacttggtga atcgtatcga actgaagggc atcgacttca aagaggacgg caacattctg 420  
 ggccacaagc tggagtacaa cttcaacagc cactacgtgt acatcaccgc cgataagcag 480  
 aagaactgca tcaaggccaa cttcaagatt cgtcacaacg tggaggatgg ctccgtgcag 540  
 ctggccgatc actaccagca gaacacaccc atcggcgatg gaccggtttt actgcccgac 600  
 aaccactatt taagccacca gagcaagctg tccaaggacc ccaacgagaa gcgtgatcat 660  
 atggtgctgc tcgagtttgt gaccgccgcc ggcatcacc atggaatgga cgagctgtac 720  
 aagagccggc tccatatggg atccggcgga ctcagtagcc ccaagaagaa acgcaaagtc 780  
 gaggatccaa agaagaaaag gaaggttgaa gacccaaga aaaagaggaa ggtgggttcc 840  
 gactataagg accacgacgg agactacaag gatcatgata ttgattacaa agacgatgac 900  
 gataagatgg ccccaaagaa gaagcggaag gtcggacgct agmagvskgt gvvvdgdvng 960  
 hksvrgggda tngktkcttg kvwtvttgyg vacsrydhmk hdksamgyvr tskddgtykt 1020

|                                                                    |      |
|--------------------------------------------------------------------|------|
| ravkgdtnr kgdkdgnghk ynnshvyta dkknckankr hnvdgsvadh yntgdgvdnh    | 1080 |
| yshskskdnk rdhmvvtaag thgm dyksrh mgsggsskkk rkvdkkkrkv dkkkrkvgsd | 1140 |
| ykdhgdgdykd hddykdddkd makkkkrkvgr                                 | 1170 |

<210> 37  
 <211> 862  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> NL control

<220>  
 <221> misc\_feature  
 <222> (674)..(674)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (730)..(730)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (742)..(742)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (750)..(750)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (755)..(755)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (757)..(757)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (760)..(760)  
 <223> n is a, c, g, t or u

<220>

<221> misc\_feature  
 <222> (768)..(768)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (772)..(772)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (776)..(776)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (785)..(785)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (792)..(792)  
 <223> n is a, c, g, t or u

<220>  
 <221> misc\_feature  
 <222> (802)..(802)  
 <223> n is a, c, g, t or u

<400> 37  
 atggccggag tgttcacact ggaggacttc gtgggcgact ggagacagac cgccggctac 60  
 aacctggacc aggtgctgga gcagggcgga gtgagcagcc tgtttcagaa cctgggctgt 120  
 agcgtcacc ccatccagag gatcgtgctg tccggcgaga atggcctgaa gatcgacatc 180  
 cacgtcatca tcccctacga gggcctgagc ggcgatcaga tgggccagat cgagaagatc 240  
 ttcaaggtgg tgtatccgt cgacgaccac cacttcaagg tgatcctgca ttacggcacc 300  
 ctctgtgatcg acggcgtgac ccctaacaatg atcgactact tcggcaggcc ctacgaggga 360  
 atcgccgtgt tcgacggaaa gaagatcacc gtgaccggca ccctgtggaa cggaaacaag 420  
 atcatcgacg agaggctgat caaccccgac ggctccctgc tgttcagggt gaccatcaat 480  
 ggcgtgaccg gctggagact gtgcgagaga atcctggccg gaggcggagg aagcctcgtg 540  
 cccagaggat ccggcggagg cggctccagc cggctccata tgggatccgg cggactcagt 600  
 agcggctccg gcggcagcgg gggaggcggg catcatcacc atcaccacgg acgctagmag 660  
 vskgtgvvvd gdvnghksvs gggdatygkt kcttgkvwtv tttgvcry dhmkhdksam 720

gyvrtkddgn yktravkgdt vnrkgdkdgn ghkynynshn vymadkkngk vnkrhndgsv 780  
 adhyntgdgv dnhystsask dnkrdhmvvt aagtgmddykk kkrkvshgsg gsgggsgdyk 840  
 dhdgdykdhd dykdddkma gr 862

<210> 38  
 <211> 30  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> RNA probe

<400> 38  
 gacaugccag cagcagcagc agcaggacug 30

<210> 39  
 <211> 30  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> RNA probe

<400> 39  
 gacaugccgg cggcggcggc ggccgggacug 30

<210> 40  
 <211> 30  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> RNA probe

<400> 40  
 gacacugcug cugcugcugc ugcugaugca 30

<210> 41  
 <211> 30  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> RNA probe

<400> 41

gacaccgccg ccgccgccgc cgccggacug 30

<210> 42  
 <211> 30  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> RNA probe

<400> 42  
 gacaugcugg uguaucuugu cuuuagacug 30

<210> 43  
 <211> 63  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-a

<400> 43  
 gaagacataa actccgtggc cacatacaga gaccaaggc tcagtggc catacatg 60

ttc 63

<210> 44  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-b

<400> 44  
 gaagacatat acagagacca aggtctcag ggtgacataa tgtcttc 47

<210> 45  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-c

<400> 45  
 gaagacatca tacagagacc aaggtctcag tggttacata tgtcttc 47

<210> 46  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-d

<400> 46  
 gaagacatac atacagagac caaggtctca gtggttaca tgtcttc 47

<210> 47  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-e

<400> 47  
 gaagacatta catacagaga ccaaggtctc agtggtgaca tgtcttc 47

<210> 48  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-f

<400> 48  
 gaagacattg acatacagag accaaggtct cagtgggtaa tgtcttc 47

<210> 49  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> Dest-g

<400> 49  
 gaagacatgt tacatacaga gaccaaggtc tcagtgggtca tgtcttc 47

<210> 50  
 <211> 47  
 <212> DNA  
 <213> Artificial sequence

&lt;220&gt;

&lt;223&gt; Dest-h

&lt;400&gt; 50

gaagacatgg tcacatacag agaccaaggt ctcatggta tgtcttc

47

&lt;210&gt; 51

&lt;211&gt; 47

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Dest-i

&lt;400&gt; 51

gaagacattg gttacataca gagaccaagg tctcatgga tgtcttc

47

&lt;210&gt; 52

&lt;211&gt; 45

&lt;212&gt; DNA

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; Dest-j

&lt;400&gt; 52

gaagacatgt ggtgacatac agagaccaag gtctcatgg tcttc

45

&lt;210&gt; 53

&lt;211&gt; 5

&lt;212&gt; PRT

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; signal sequence

&lt;400&gt; 53

Met Gly Asn Ser Val

1

5

&lt;210&gt; 54

&lt;211&gt; 22

&lt;212&gt; PRT

&lt;213&gt; Artificial sequence

&lt;220&gt;

&lt;223&gt; signal sequence

&lt;400&gt; 54

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
|     |     |     |     |     |     |     |     |     |     | 118 |     |     |     |     |     |  |  |  |  |
| Glu | Leu | Thr | Tyr | Asn | Thr | Leu | Ile | Ser | Gly | Leu | Gly | Lys | Ala | Gly | Arg |  |  |  |  |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |  |  |  |  |

Ala Arg Asp Pro Pro Val  
20

<210> 55  
<211> 4  
<212> PRT  
<213> Artificial sequence

<220>  
<223> signal sequence

<400> 55  
Lys Asp Glu Leu  
1

<210> 56  
<211> 4  
<212> PRT  
<213> Artificial sequence

<220>  
<223> signal sequence

<400> 56  
Lys Glu Glu Leu  
1

<210> 57  
<211> 19  
<212> PRT  
<213> Artificial sequence

<220>  
<223> signal sequence

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |  |  |  |
| Met | Gly | Trp | Ser | Cys | Ile | Ile | Leu | Phe | Leu | Val | Ala | Thr | Ala | Thr | Gly |  |  |  |  |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |  |  |  |  |

Ala His Ser

<210> 58  
<211> 35  
<212> PRT

<213> Artificial Sequence

<220>

<223> 1st\_A (v3.2\_A)

<400> 58

Val Thr Tyr Thr Thr Asn Ile Asp Gln Leu Cys Lys Ala Gly Lys Val  
1 5 10 15

Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys  
20 25 30

Pro Asn Val  
35

<210> 59

<211> 35

<212> PRT

<213> Artificial Sequence

<220>

<223> 1st\_G (v3.2\_G)

<400> 59

Val Thr Tyr Thr Thr Asn Ile Asp Gln Leu Cys Lys Ala Gly Lys Val  
1 5 10 15

Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys  
20 25 30

Pro Asp Val  
35

<210> 60

<211> 35

<212> PRT

<213> Artificial Sequence

<220>

<223> 1st\_U (v3.2\_U)

<400> 60

Val Thr Tyr Asn Thr Asn Ile Asp Gln Leu Cys Lys Ala Gly Arg Leu  
1 5 10 15

Asp Glu Ala Glu Glu Leu Leu Glu Glu Met Glu Glu Lys Gly Ile Lys  
20 25 30

Pro Asp Val  
35

<210> 61  
<211> 2271  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> v2\_polypucleotide

<400> 61  
atgggtcatc accatcatca tcacgggtcc ctgcaggact cagaagtcaa tcaagaagct 60  
aagccagagg tcaagccaga agtcaagcct gagactcaca tcaatttaaa ggtgtccgat 120  
ggatcttcag agatcttctt caagatcaaa aagaccactc ctttaagaag gctgatggaa 180  
gcgttcgcta aaagacaggg taaggaaatg gactccttaa gattcttgta cgacggtatt 240  
agaattcaag ctgatcaggc ccctgaagat ttggacatgg aggataacga tattattgag 300  
gctcaccgcg aacagattgg aggtatggga aactccgtgg tcacatacaa caccctgatac 360  
gacggcctgt gcaagagcgg caagatcgag gaggccctga agctgttcaa ggagatggag 420  
gagaagggca tcacccccag cgtgggtcaca tacaccacac tgatcgacgg actgtgtaaa 480  
gccggcgacg tggacgaagc cctcgagctg ttcaaagaga tgcggagcaa gggcgtgaag 540  
cccaacgtgg tgacatacac caccctgatac gacggcctgt gcaaggccgg caaagtggac 600  
gaggccctgg agctgttcga cgagatgaag gagaggggca tcaagcccga cgtgggtcaca 660  
tacaacaccc tgatcgacgg cctgtgcaag agcggcaaga tcgaggaggc cctgaagctg 720  
ttcaaggaga tggaggagaa gggcatcacc ccagcgtgg ttacatacac cacactgatac 780  
gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg 840  
agcaagggcg tgaagcccaa cgtgggtcaca tacaccaccc tgatcgacgg cctgtgcaag 900  
gccggcaagg tggatgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag 960  
cccagcgtgg ttacatacaa caccctgatac gacggcctgt gcaagagcgg caagatcgag 1020  
gaggccctga agctgttcaa ggagatggag gagaagggca tcacccccag cgtgggtcaca 1080  
tacaccacac tgatcgacgg actgtgtaaa gccggcgacg tggacgaagc cctcgagctg 1140

```

ttcaaagaga tgcggagcaa gggcgtgaag cccaacgtgg tgacatacac caccctgatac 1200
gacggcctgt gcaaggccgg caaagtggac gaggccctgg agctgttcga cgagatgaag 1260
gagaggggca tcaagcccga cgtgggtcaca tacaacaccc tgatcgacgg cctgtgcaag 1320
agcggcaaga tcgaggaggc cctgaagctg ttcaaggaga tggaggagaa gggcatcacc 1380
cccagcgtgg ttacatacac cacactgatac gacggactgt gtaaagccgg cgacgtggac 1440
gaagccctcg agctgttcaa agagatgcgg agcaagggcg tgaagcccaa cgtgggtcaca 1500
tacaccaccc tgatcgacgg cctgtgcaag gccggcaagg tggatgaggc cctggagctg 1560
ttcgacgaga tgaaggagag gggcatcaag cccgacgtgg tcacatacaa caccctgatac 1620
gacggcctgt gcaagagcgg caagatcgag gaggccctga agctgttcaa ggagatggag 1680
gagaagggca tcacccccag cgtgggtcaca tacaccacac tgatcgacgg actgtgtaaa 1740
gccggcgacg tggacgaagc cctcgagctg ttcaaagaga tgcggagcaa gggcgtgaag 1800
cccaacgtgg ttacatacac caccctgatac gacggcctgt gcaaggccgg caaagtggac 1860
gaggccctgg agctgttcga cgagatgaag gagaggggca tcaagcccga cgtgggtcaca 1920
tacaacaccc tgatcgacgg cctgtgcaag agcggcaaga tcgaggaggc cctgaagctg 1980
ttcaaggaga tggaggagaa gggcatcacc cccagcgtgg tgacatacac cacactgatac 2040
gacggactgt gtaaagccgg cgacgtggac gaagccctcg agctgttcaa agagatgcgg 2100
agcaagggcg tgaagcccaa cgtgggtcaca tacaccaccc tgatcgacgg cctgtgcaag 2160
gccggcaagg tggatgaggc cctggagctg ttcgacgaga tgaaggagag gggcatcaag 2220
cccgacgagc tgacctacaa caccctgatac agcggcctgg gcaaggccgg c 2271

```

<210> 62  
 <211> 757  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> v2\_protein

<400> 62

Met Gly His His His His His His Gly Ser Leu Gln Asp Ser Glu Val  
 1 5 10 15

122  
 Asn Gln Glu Ala Lys Pro Glu Val Lys Pro Glu Val Lys Pro Glu Thr  
 20 25 30  
 His Ile Asn Leu Lys Val Ser Asp Gly Ser Ser Glu Ile Phe Phe Lys  
 35 40 45  
 Ile Lys Lys Thr Thr Pro Leu Arg Arg Leu Met Glu Ala Phe Ala Lys  
 50 55 60  
 Arg Gln Gly Lys Glu Met Asp Ser Leu Arg Phe Leu Tyr Asp Gly Ile  
 65 70 75 80  
 Arg Ile Gln Ala Asp Gln Ala Pro Glu Asp Leu Asp Met Glu Asp Asn  
 85 90 95  
 Asp Ile Ile Glu Ala His Arg Glu Gln Ile Gly Gly Met Gly Asn Ser  
 100 105 110  
 Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys  
 115 120 125  
 Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile  
 130 135 140  
 Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys  
 145 150 155 160  
 Ala Gly Asp Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser  
 165 170 175  
 Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile Asp Gly  
 180 185 190  
 Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu  
 195 200 205  
 Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu  
 210 215 220  
 Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu

|            |            |            |            |            |            |            |            |            |            |            |            |            |            |            |            |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
|            |            |            |            |            |            |            |            | 123        |            |            |            |            |            |            |            |
| 225        |            |            |            | 230        |            |            |            | 235        |            |            |            | 240        |            |            |            |
| Phe        | Lys        | Glu        | Met        | Glu<br>245 | Glu        | Lys        | Gly        | Ile        | Thr<br>250 | Pro        | Ser        | Val        | Val        | Thr<br>255 | Tyr        |
| Thr        | Thr        | Leu        | Ile<br>260 | Asp        | Gly        | Leu        | Cys        | Lys<br>265 | Ala        | Gly        | Asp        | Val        | Asp<br>270 | Glu        | Ala        |
| Leu        | Glu        | Leu<br>275 | Phe        | Lys        | Glu        | Met        | Arg<br>280 | Ser        | Lys        | Gly        | Val        | Lys<br>285 | Pro        | Asn        | Val        |
| Val        | Thr<br>290 | Tyr        | Thr        | Thr        | Leu        | Ile<br>295 | Asp        | Gly        | Leu        | Cys        | Lys<br>300 | Ala        | Gly        | Lys        | Val        |
| Asp<br>305 | Glu        | Ala        | Leu        | Glu        | Leu<br>310 | Phe        | Asp        | Glu        | Met        | Lys<br>315 | Glu        | Arg        | Gly        | Ile        | Lys<br>320 |
| Pro        | Asp        | Val        | Val        | Thr<br>325 | Tyr        | Asn        | Thr        | Leu        | Ile<br>330 | Asp        | Gly        | Leu        | Cys        | Lys<br>335 | Ser        |
| Gly        | Lys        | Ile        | Glu<br>340 | Glu        | Ala        | Leu        | Lys        | Leu<br>345 | Phe        | Lys        | Glu        | Met        | Glu<br>350 | Glu        | Lys        |
| Gly        | Ile<br>355 | Thr        | Pro        | Ser        | Val        | Val        | Thr<br>360 | Tyr        | Thr        | Thr        | Leu        | Ile<br>365 | Asp        | Gly        | Leu        |
| Cys<br>370 | Lys        | Ala        | Gly        | Asp        | Val        | Asp<br>375 | Glu        | Ala        | Leu        | Glu        | Leu<br>380 | Phe        | Lys        | Glu        | Met        |
| Arg<br>385 | Ser        | Lys        | Gly        | Val        | Lys<br>390 | Pro        | Asn        | Val        | Val<br>395 | Thr        | Tyr        | Thr        | Thr        | Leu        | Ile<br>400 |
| Asp        | Gly        | Leu        | Cys<br>405 | Lys        | Ala        | Gly        | Lys        | Val        | Asp<br>410 | Glu        | Ala        | Leu        | Glu        | Leu<br>415 | Phe        |
| Asp        | Glu        | Met        | Lys<br>420 | Glu        | Arg        | Gly        | Ile        | Lys<br>425 | Pro        | Asp        | Val        | Val        | Thr<br>430 | Tyr        | Asn        |
| Thr        | Leu        | Ile<br>435 | Asp        | Gly        | Leu        | Cys        | Lys<br>440 | Ser        | Gly        | Lys        | Ile        | Glu<br>445 | Glu        | Ala        | Leu        |

Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val Val  
 450 455 460

Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Asp Val Asp  
 465 470 475 480

Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro  
 485 490 495

Asn Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly  
 500 505 510

Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly  
 515 520 525

Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys  
 530 535 540

Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu  
 545 550 555 560

Glu Lys Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp  
 565 570 575

Gly Leu Cys Lys Ala Gly Asp Val Asp Glu Ala Leu Glu Leu Phe Lys  
 580 585 590

Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr  
 595 600 605

Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu  
 610 615 620

Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr  
 625 630 635 640

Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu  
 645 650 655

Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser  
                   660                  665                  670

Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Asp  
           675                  680                  685

Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val  
       690                  695                  700

Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys  
   705                  710                  715                  720

Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu  
                   725                  730                  735

Arg Gly Ile Lys Pro Asp Glu Leu Thr Tyr Asn Thr Leu Ile Ser Gly  
           740                  745                  750

Leu Gly Lys Ala Gly  
           755

<210> 63  
 <211> 2274  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> v3.2\_polypucleotide

<400> 63  
 atgggtcatc accatcatca tcacgggtcc ctgcaggact cagaagtcaa tcaagaagct 60  
 aagccagagg tcaagccaga agtcaagcct gagactcaca tcaatttaaa ggtgtccgat 120  
 ggatcttcag agatcttctt caagatcaaa aagaccactc cttaagaag gctgatggaa 180  
 gcgttcgcta aaagacaggg taaggaaatg gactccttaa gattcttgta cgacgttatt 240  
 agaattcaag ctgatcaggc ccctgaagat ttggacatgg aggataacga tattattgag 300  
 gctcaccgcg aacagattgg aggtatggga aactccgtgg tcacatacaa caccaacatc 360  
 gaccagctgt gcaagagcgg caagatcgag gaggccctga agctgttcaa ggagatggag 420  
 gagaagggca tcacccccag cgtgggtcaca tacaccacac tgatcgacgg actgtgtaaa 480

|            |            |             |            |            |             |      |
|------------|------------|-------------|------------|------------|-------------|------|
| gccggcgacg | tggacgaagc | cctcgagctg  | ttcaaagaga | tgcggagcaa | gggcgtgaag  | 540  |
| cccaacgtgg | tgacatacac | caccctgatac | gacggcctgt | gcaaggccgg | caaagtggac  | 600  |
| gaggccctgg | agctgttcga | cgagatgaag  | gagaggggca | tcaagcccga | cgtggtcaca  | 660  |
| tacaacaccc | tgatcgacgg | cctgtgcaag  | agcggcaaga | tcgaggaggc | cctgaagctg  | 720  |
| ttcaaggaga | tggaggagaa | gggcatcacc  | cccagcgtgg | ttacatacac | cacactgatac | 780  |
| gacggactgt | gtaaagccgg | cgacgtggac  | gaagccctcg | agctgttcaa | agagatgcgg  | 840  |
| agcaagggcg | tgaagcccaa | cgtggtcaca  | tacaccaccc | tgatcgacgg | cctgtgcaag  | 900  |
| gccggcaagg | tggatgaggc | cctggagctg  | ttcgacgaga | tgaaggagag | gggcatcaag  | 960  |
| cccgacgtgg | ttacatacaa | caccctgatac | gacggcctgt | gcaagagcgg | caagatcgag  | 1020 |
| gaggccctga | agctgttcaa | ggagatggag  | gagaagggca | tcacccccag | cgtggtcaca  | 1080 |
| tacaccacac | tgatcgacgg | actgtgtaaa  | gccggcgacg | tggacgaagc | cctcgagctg  | 1140 |
| ttcaaagaga | tgcggagcaa | gggcgtgaag  | cccaacgtgg | tgacatacac | caccctgatac | 1200 |
| gacggcctgt | gcaaggccgg | caaagtggac  | gaggccctgg | agctgttcga | cgagatgaag  | 1260 |
| gagaggggca | tcaagcccga | cgtggtcaca  | tacaacaccc | tgatcgacgg | cctgtgcaag  | 1320 |
| agcggcaaga | tcgaggaggc | cctgaagctg  | ttcaaggaga | tggaggagaa | gggcatcacc  | 1380 |
| cccagcgtgg | ttacatacac | cacactgatac | gacggactgt | gtaaagccgg | cgacgtggac  | 1440 |
| gaagccctcg | agctgttcaa | agagatgcgg  | agcaagggcg | tgaagcccaa | cgtggtcaca  | 1500 |
| tacaccaccc | tgatcgacgg | cctgtgcaag  | gccggcaagg | tggatgaggc | cctggagctg  | 1560 |
| ttcgacgaga | tgaaggagag | gggcatcaag  | cccgacgtgg | tcacatacaa | caccctgatac | 1620 |
| gacggcctgt | gcaagagcgg | caagatcgag  | gaggccctga | agctgttcaa | ggagatggag  | 1680 |
| gagaagggca | tcacccccag | cgtggtcaca  | tacaccacac | tgatcgacgg | actgtgtaaa  | 1740 |
| gccggcgacg | tggacgaagc | cctcgagctg  | ttcaaagaga | tgcggagcaa | gggcgtgaag  | 1800 |
| cccaacgtgg | ttacatacac | caccctgatac | gacggcctgt | gcaaggccgg | caaagtggac  | 1860 |
| gaggccctgg | agctgttcga | cgagatgaag  | gagaggggca | tcaagcccga | cgtggtcaca  | 1920 |
| tacaacaccc | tgatcgacgg | cctgtgcaag  | agcggcaaga | tcgaggaggc | cctgaagctg  | 1980 |
| ttcaaggaga | tggaggagaa | gggcatcacc  | cccagcgtgg | tgacatacac | cacactgatac | 2040 |

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| 127        |            |            |            |            |            |      |
| gacggactgt | gtaaagccgg | cgacgtggac | gaagccctcg | agctgttcaa | agagatgcgg | 2100 |
| agcaagggcg | tgaagcccaa | cgtggtcaca | tacaccaccc | tgatcgacgg | cctgtgcaag | 2160 |
| gccggcaagg | tggatgaggc | cctggagctg | ttcgacgaga | tgaaggagag | gggcatcaag | 2220 |
| cccgacgagg | agctgacctg | caacaccctg | atcagcggcc | tgggcaaggc | cggc       | 2274 |

|       |                     |
|-------|---------------------|
| <210> | 64                  |
| <211> | 757                 |
| <212> | PRT                 |
| <213> | Artificial Sequence |

<211> 757

&lt;212&gt; PRT

### <213> Artificial Sequence

 $\langle 220 \rangle$ 

<223> v3.2\_protein

<400> 64

Met Gly His His His His His His Gly Ser Leu Gln Asp Ser Glu Val  
1 5 10 15

Asn Gln Glu Ala Lys Pro Glu Val Lys Pro Glu Val Lys Pro Glu Thr

His Ile Asn Leu Lys Val Ser Asp Gly Ser Ser Glu Ile Phe Phe Lys  
35 40 45

Ile Lys Lys Thr Thr Pro Leu Arg Arg Leu Met Glu Ala Phe Ala Lys  
50 55 60

Arg Gln Gly Lys Glu Met Asp Ser Leu Arg Phe Leu Tyr Asp Gly Ile  
65 70 75 80

Arg Ile Gln Ala Asp Gln Ala Pro Glu Asp Leu Asp Met Glu Asp Asn  
85 90 95

Asp Ile Ile Glu Ala His Arg Glu Gln Ile Gly Gly Met Gly Asn Ser  
100 105 110

Val Val Thr Tyr Asn Thr Asn Ile Asp Gln Leu Cys Lys Ser Gly Lys  
115 120 125

Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile  
130 135 140

Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys  
145 150 155 160

Ala Gly Asp Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser  
165 170 175

Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile Asp Gly  
180 185 190

Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu  
195 200 205

Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu  
210 215 220

Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu  
225 230 235 240

Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val Val Thr Tyr  
245 250 255

Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Asp Val Asp Glu Ala  
260 265 270

Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro Asn Val  
275 280 285

Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Lys Val  
290 295 300

Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly Ile Lys  
305 310 315 320

Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys Lys Ser  
325 330 335

Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu Glu Lys  
340 345 350

Gly Ile Thr Pro Ser Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu  
 355 360 365

Cys Lys Ala Gly Asp Val Asp Glu Ala Leu Glu Leu Phe Lys Glu Met  
 370 375 380

Arg Ser Lys Gly Val Lys Pro Asn Val Val Thr Tyr Thr Thr Leu Ile  
 385 390 395 400

Asp Gly Leu Cys Lys Ala Gly Lys Val Asp Glu Ala Leu Glu Leu Phe  
 405 410 415

Asp Glu Met Lys Glu Arg Gly Ile Lys Pro Asp Val Val Thr Tyr Asn  
 420 425 430

Thr Leu Ile Asp Gly Leu Cys Lys Ser Gly Lys Ile Glu Glu Ala Leu  
 435 440 445

Lys Leu Phe Lys Glu Met Glu Glu Lys Gly Ile Thr Pro Ser Val Val  
 450 455 460

Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly Asp Val Asp  
 465 470 475 480

Glu Ala Leu Glu Leu Phe Lys Glu Met Arg Ser Lys Gly Val Lys Pro  
 485 490 495

Asn Val Val Thr Tyr Thr Thr Leu Ile Asp Gly Leu Cys Lys Ala Gly  
 500 505 510

Lys Val Asp Glu Ala Leu Glu Leu Phe Asp Glu Met Lys Glu Arg Gly  
 515 520 525

Ile Lys Pro Asp Val Val Thr Tyr Asn Thr Leu Ile Asp Gly Leu Cys  
 530 535 540

Lys Ser Gly Lys Ile Glu Glu Ala Leu Lys Leu Phe Lys Glu Met Glu  
 545 550 555 560

[illegible]