SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION

Inventors: Keith V. Wood, Madison, WI (US); Monika G. Gruber, Madison, WI (US); Yao Zhuang, Madison, WI (US); Aileen Pargno, Madison, WI (US)

Correspondence Address:
Schwagman, Lundberg, Woessner & Kluth, P.A.
P.O. Box 2938
Minneapolis, MN 55402 (US)

Assignee: Promega Corporation

Related U.S. Application Data
Division of application No. 09/645,706, filed on Aug. 24, 2000.

Publication Classification
Int.Cl.  
C12P 19/34 (2006.01)  
C07H 21/04 (2006.01)  
C12N 15/63 (2006.01)

U.S. Cl. ................ 435/91.1; 536/23.2; 435/320.1

ABSTRACT
A method to prepare synthetic nucleic acid molecules having reduced inappropriate or unintended transcriptional characteristics when expressed in a particular host cell.
## Figure 1

The Genetic Code

<table>
<thead>
<tr>
<th>First Position (5' end)</th>
<th>Second position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U</td>
<td>C</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Stop</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Met</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Ala</td>
</tr>
</tbody>
</table>
### Figure 2 (cont.)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
</table>
Figure 2 (cont.)

```
<table>
<thead>
<tr>
<th>Sequence</th>
<th>GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRV611.1</td>
<td>AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG</td>
</tr>
</tbody>
</table>

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

YGB1-6G1.1 | GCATCGCTACCTCTGGACGTTATTCTGAGGAGAATTGATTAAAG |

RDOV1.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV2.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV3.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV4.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
RDOV5.1    | AATTTTACGCTGCTGGACGTTATTCTGAGGAGAATTGATTAAAG |
```

Figure 3 (cont.)

GRVER51.SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GR6. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GRVER5. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GRVER4. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GRVER3. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GRVER2. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
GRVER1. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
YG81-6G1. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER1. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER2. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER3. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER4. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER5. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RO1. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER1. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER2. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER3. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER4. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDVER5. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558
RDS16H9. SEQ FVVKQPGKEITAKEYDYLAEVRSHKTLYRGGVRFVDSP1558

GRVER51.SEQ RNVGTKTRKELKQLLEKAGG 1624
GR6. SEQ RNVGTKTRKELKQLLEKAGG 1624
GRVER5. SEQ RNVGTKTRKELKQLLEKAGG 1624
GRVER4. SEQ RNVGTKTRKELKQLLEKAGG 1624
GRVER3. SEQ RNVGTKTRKELKQLLEKAGG 1624
GRVER2. SEQ RNVGTKTRKELKQLLEKAGG 1624
GRVER1. SEQ RNVGTKTRKELKQLLEKAGG 1624
YG81-6G1. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER1. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER2. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER3. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER4. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER5. SEQ RNVGTKTRKELKQLLEKAGG 1624
RO1. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER1. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER2. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER3. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER4. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDVER5. SEQ RNVGTKTRKELKQLLEKAGG 1624
RDS16H9. SEQ RNVGTKTRKELKQLLE[1]KAGG 1624
Figure 4 Codon Usage Analysis

<table>
<thead>
<tr>
<th>Codon</th>
<th>YGGR±4G</th>
<th>YGGR</th>
<th>YGGR</th>
<th>YGGR</th>
<th>YGGR</th>
<th>YGGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CTA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ACA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ACC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CCC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CGC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The table above shows the codon usage analysis per seat epides free. The table compares the usage of different codons across various sequences, indicating which codons are more frequently used in the sequenced samples. The relative codon usage for each seq (*1000) is also provided for comparison.
### Figure 5A

**Codon Usage YG#81-6G01 (yellow-green)**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
</tr>
<tr>
<td>CGG</td>
<td>Pro</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
</tr>
<tr>
<td>TAA</td>
<td>Tyr</td>
</tr>
<tr>
<td>TAG</td>
<td>Tyr</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
</tr>
<tr>
<td>TGA</td>
<td>Cys</td>
</tr>
<tr>
<td>TGG</td>
<td>Trp</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
</tr>
</tbody>
</table>
**Figure 5B**

**Codon Usage: GRver1**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>16</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>9</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>10</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>28</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CGG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>6</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>7</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>7</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>17</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>18</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>19</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>19</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>19</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>19</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>13</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>19</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>
**Figure 5C**

**Codon Usage: RDver1**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>20</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>19</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>24</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>15</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>19</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>10</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>0</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>14</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>0</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>14</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>0</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>18</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>0</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>19</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>19</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>6</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>5</td>
</tr>
<tr>
<td>TGA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>13</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>13</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CAG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>15</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AAG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Gly</td>
<td>20</td>
</tr>
<tr>
<td>GCC</td>
<td>Gly</td>
<td>19</td>
</tr>
<tr>
<td>GCA</td>
<td>Gly</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>

**Notes:**
- ATG, TGA, TAA, TAT, TAG, TGC, TGT, TCT, TCC, TCA, TCG, TTA, CTT, CTC, CTA, and CCG codons are non-coding.
**Figure 5D**

**Codon Usage: Grver2**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>28</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>20</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>18</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>28</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>15</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>14</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>19</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>9</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>10</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>6</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>10</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>4</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>16</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>19</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>17</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>5</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>6</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>13</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>13</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>16</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>18</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>21</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>0</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>
### Codon Usage: Rdver2

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>19</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>20</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>21</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>28</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>16</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>0</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>15</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>13</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>11</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>19</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>17</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>1</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>13</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>21</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>17</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>21</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>13</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>0</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>

Counts:
- Phe: 6
- Cys: 5
- Tyr: 10
- Ser: 21
- Glu: 7
- His: 13
- Arg: 13
- Thr: 11
- Ala: 19
- Asp: 13
- Glu: 21
- Val: 0
- Gly: 18
Figure 5F

Codon Usage: GRver3

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>26</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>5</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>24</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>14</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>24</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>1</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>22</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>27</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>16</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>18</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>10</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>14</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>8</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>1</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>12</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>14</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>6</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>9</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>5</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>21</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>14</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>12</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>14</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>20</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>18</td>
</tr>
</tbody>
</table>

- TAA and TGA are stop codons.
- TGT is Cys.
- TGC is Cys.
- TGG is Trp.

The table above shows the codon usage for different amino acids, with respective frequencies.
### Figure 5G

**Codon Usage: RDver3**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>27</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>14</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>7</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>13</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>6</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>22</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>16</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>12</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>10</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>3</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>8</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>7</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>20</td>
</tr>
<tr>
<td>AIC</td>
<td>Ile</td>
<td>19</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTC</td>
<td>Ala</td>
<td>20</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>16</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>1</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>14</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>12</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>16</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>10</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>15</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>16</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>23</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>0</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>
## Figure 5H

**Codon Usage: GRver4**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>11</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>14</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Leu</td>
<td>21</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>11</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>22</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>13</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>25</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>3</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>22</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>13</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>TGG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>18</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>10</td>
</tr>
<tr>
<td>CGG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>14</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>8</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>19</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>15</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>7</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>12</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>7</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>6</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>11</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>3</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>20</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>15</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>12</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>14</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>14</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>11</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>1</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>14</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>17</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>19</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>3</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>
### Figure 51

**Codon Usage: RDver4**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>28</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>7</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>2</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>7</td>
</tr>
<tr>
<td>TAG</td>
<td>Tyr</td>
<td>0</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>4</td>
</tr>
<tr>
<td>TGG</td>
<td>Trp</td>
<td>2</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>11</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>15</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>2</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>11</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>10</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>7</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>8</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>13</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>22</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>15</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>18</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>20</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>22</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 5J

Codon Usage: GRver5

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>10</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>15</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>23</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>12</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>19</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>15</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>23</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>3</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>21</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>1</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>25</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>11</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>4</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>17</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>2</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>9</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>14</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>8</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>ACG</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>18</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>14</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>5</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>7</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>12</td>
</tr>
<tr>
<td>TAA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Trp</td>
<td>2</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>6</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>11</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>3</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>9</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>13</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>19</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>16</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>12</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>14</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>19</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>19</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>8</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>3</td>
</tr>
<tr>
<td>TGA</td>
<td>***</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Trp</td>
<td>2</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>13</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>11</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>2</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>14</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>16</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>21</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>1</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>1</td>
</tr>
</tbody>
</table>
**Figure 5K**

**Codon Usage: RDver5**

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>13</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>25</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>11</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>2</td>
</tr>
<tr>
<td>TGG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CTA</td>
<td>Pro</td>
<td>15</td>
</tr>
<tr>
<td>CGG</td>
<td>Pro</td>
<td>16</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>19</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>20</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>11</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>5</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>26</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>1</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>17</td>
</tr>
</tbody>
</table>

**Amino Acids:**
- Phe: Phenylalanine
- Ser: Serine
- Tyr: Tyrosine
- Cys: Cysteine
- Ala: Alanine
- Arg: Arginine
- His: Histidine
- Asn: Asparagine
- Glu: Glutamic Acid
- Lys: Lysine
- Thr: Threonine
- Gln: Glutamine
- Trp: Tryptophan

**Counts:**
- Phe: 7
- Ser: 2
- Tyr: 13
- Cys: 4
- Ala: 21
- Arg: 0
- His: 4
- Asn: 9
- Glu: 18
- Lys: 13
- Thr: 10
- Gln: 8
- Trp: 2
- Asp: 14
- Gly: 14
- Val: 26
- Gln: 7
- Asp: 12
- Gly: 14
- Val: 17
- Gln: 0
- Asp: 0
- Gly: 1
Figure 6

Synthetic oligos for engineered GR/RD genes
(All oligos listed 5' to 3')

Coding strand: 5'______ (__)n______ 3'
Non-coding strand: 3'______ (__)n______ 5'

Oligos with pRAM flanking sequence identical for GR/RD
1) coding strand upstream flanking
   RAM-C1: ACGCCAGCCCAAGCCTAGGCCTGAGTGC (SEQ ID NO:35)
   RAM-C2: CTTAAATTCTCCCATCCCCCTGTTGACAAATATACCGTGCTG (SEQ ID NO:36)
   RAM-C3: TATAATGTTAGGAATTTGCCAGGGATAACAATTTCAACA (SEQ ID NO:37)

2) coding strand downstream flanking
   RAM-C4: ATGGGATTTTGGCTACCATAGCAATGGAATTTTGGTAAT (SEQ ID NO:38)
   RAM-C5: AAGTGGTTAATGGATATTGCTGAAATCCTGGGTTTTTTTTT (SEQ ID NO:39)
   RAM-C6: GATATCAAGCTTATCGAGTCCGACCTCGAGGATTATA (SEQ ID NO:40)
   RAM-C7: TAGAAAGGGCTCGGGGCGGGGCCTGAGTGGCGTT (SEQ ID NO:41)

3) non-coding strand downstream flanking
   RAM-N1: AACACTGACTTGGACGTCCAAAG (SEQ ID NO:42)
   RAM-N2: GCCGCCAGCCCAAGCCTAGGCCTGAGTGC (SEQ ID NO:43)
   RAM-N3: GTATCGTATAGAGCTTGATATCGGAAATCTCGGTT (SEQ ID NO:44)
   RAM-N3b: CTGGTGGATATGGATAATCGGAAATCTCGGTT (SEQ ID NO:45)
   RAM-N4: TGGAAAAATTCATTTAGCAAGGATTTTCTACCAATTTTCATAT (SEQ ID NO:46)
   RAM-N5: TGGTCTAGTAACTCCCTCACATCAGTCTGTTTTTTTATA (SEQ ID NO:47)

4) non-coding strand upstream flanking
   RAM-N6: TGGCAATTCTCCCTACCTTACATGAGGATATATGCTGCTG (SEQ ID NO:48)
   RAM-N7: AACAGGGGGGAGAAATGGGAAACTGCCCTAAGCTGGGCTGCGT (SEQ ID NO:49)

Grver5 with flanking seq. of pRAM to end of Sfi I primers
1) Coding strand (Start and stop codons are underlined)
   GR-C1: GGAACACAGATTCTCCATGATGACGGAGAGAAGGTGTAT (SEQ ID NO:50)
   GR-C2: CTTAGCCGAGCAGACGAGCTCTGCACTCAGTGAGGAGCCTCC (SEQ ID NO:51)
   GR-C3: GCTGGTAGAGTGGCTCTCCCGAGCAAGCCGAAACATCTGC (SEQ ID NO:52)
   GR-C4: ACCTCCCTCAAGACGCTGAGGCTGAGGAGCAGGAGG (SEQ ID NO:53)
   GR-C5: CTTCTCTCTACCAAAAGATTTTTTCCAAGGTACTTGTGGGT (SEQ ID NO:54)
   GR-C6: GCCCAAAGCCTCCCATATACTGGGGTTACAAATAGAGCAGT (SEQ ID NO:55)
   GR-C7: TTGGTAGCTCATTGCTGCTGAGGAATCTTACGGGCTTCCTTAT (SEQ ID NO:56)
   GR-C8: TGGTCTAGTAACTCCCTCACATCAGTCTGTTTTTTTATA (SEQ ID NO:57)
   GR-C9: CTTCTCTCTACCAAAAGATTTTTTCCAAGGTACTTGTGGGT (SEQ ID NO:58)
   GR-C10: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:59)
   GR-C11: AAAAACTCTTTGCTGCTGAGGAGCCTCCCTCCTCCTGCT (SEQ ID NO:60)
   GR-C12: CTTCTCTCTACCAAAAGATTTTTTCCAAGGTACTTGTGGGT (SEQ ID NO:61)
   GR-C13: ACATCCGAGGCTTGTGAAGCAGGCTCCCTCAATCTCTACCTCGG (SEQ ID NO:62)
   GR-C14: TTAAGAGCTGAGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:63)
   GR-C15: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:64)
   GR-C16: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:65)
   GR-C17: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:66)
   GR-C18: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:67)
   GR-C19: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:68)
   GR-C20: CTTGAGTGAATATGAGCTACTCAGGCAATCTCTACCAAA (SEQ ID NO:69)
Figure 6 (Cont.)

2) Non-coding strand

GR-N1: TAAACTAAGAAGAAGCTTTACTAGGCAGCCGACGTTTTTCGAGGA (SEQ ID NO: 92)
GR-N2: GATTGCTTAGCAACTACCTTATACAGTGTACATCTAAGGATAC (SEQ ID NO: 93)
GR-N3: GATGCCTGAAATAGAATGCTAGAAGAAGACGCGAGCGCGCCACG (SEQ ID NO: 94)
GR-N4: AGAGATTTACCCTTACGGGAAACCTGCGCACGAC (SEQ ID NO: 95)
GR-N5: ATACCCGCAGAGCGCAGCTTACAGTGAGTTGCTGAGG (SEQ ID NO: 96)
GR-N6: CCAAAAGCAGCTAGCAGATCGGCGCGCA (SEQ ID NO: 97)
GR-N7: ATACCCGCAGAGCGCAGCTTACAGTGAGTTGCTGAGG (SEQ ID NO: 98)
GR-N8: TTTGATCTTTAAAATCTCTTTTCTGGAGGTTAGG (SEQ ID NO: 99)
GR-N9: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 100)
GR-N10: AATGGTTGCTCTGCTATCATGATATGGGAGAGGTCGCGCTAT (SEQ ID NO: 101)
GR-N11: GGAGGAGGGTCATAGCTTACTTCTGGAAACATGGCCTC (SEQ ID NO: 102)
GR-N12: CAGAGTCAGCTGCGTCTGGAGGATGAGGCTTCCTAC (SEQ ID NO: 103)
GR-N13: ATACCCGCAGAGCGCAGCTTACAGTGAGTTGCTGAGG (SEQ ID NO: 104)
GR-N14: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 105)
GR-N15: CACCGAGCACGGCTACAGGACTAATAACCTGCTCGCAAG (SEQ ID NO: 106)
GR-N16: CTAAGTTGTTGAGGTTAGG (SEQ ID NO: 107)
GR-N17: CGACACGGAATAAGCCGAGCTTGTAAGAGCCGAC (SEQ ID NO: 108)
GR-N18: CTCCGAGCTTCTCTTTGGCAGAAAAGACGCAGCCGACG (SEQ ID NO: 109)
GR-N19: AGCTTCAGGAAGCTGCTGCTGGAGGATGAGGCTTCCTAC (SEQ ID NO: 110)
GR-N20: CACCGAGCACGGCTACAGGACTAATAACCTGCTCGCAAG (SEQ ID NO: 111)
GR-N21: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 112)
GR-N22: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 113)
GR-N23: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 114)
GR-N24: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 115)
GR-N25: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 116)
GR-N26: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 117)
GR-N27: GATGCTGCAAGGCTGGTGCTGGCAGCTGGGAG (SEQ ID NO: 118)
Figure 6 (Cont.)

GR-N28: ACTTGCTCGACCTGAGTCAAAATGCAAGGAGGCTTGAAATTAG
(SEQ ID NO:119)
GR-N29: CGATATTACATCATCAGCTGTAACAGGATGAATACGTTAGGAG
(SEQ ID NO:120)
GR-N30: GCTCTCAACAGCCGTGAGATTTTCTGACGATCCGAATA
(SEQ ID NO:121)
GR-N31: ATGATGCCTGGTGAGAAGTGTAGTACGAGACTGGAACCTCA
(SEQ ID NO:122)
GR-N32: AGACCTTATACAGATTTGAATGTGGATAGAAGAGATTTG
(SEQ ID NO:123)
GR-N33: AGTTTTGCTTACATCCATACCTTACACAGCTATCTGGG
(SEQ ID NO:124)
GR-N34: ATGTAAGATTCTTACACAGGAGGCGAAGCATTACATGCGAGT
(SEQ ID NO:125)
GR-N35: ACCAAGCAGCGATACAGGAATAAAAGAGCTGAGTATTAT
(SEQ ID NO:126)
GR-N36: CTCAGACAATGCTCAACACATGCTTACATTGGTGACCA
(SEQ ID NO:127)
GR-N37: CAATTATGGAGCTTTGGCCCAACAGCACGATGTCCCTGGC
(SEQ ID NO:128)
GR-N38: AAAATTTCTTGTAGAGAAGGCTCCTCCTCCTCCAGCACCTC
(SEQ ID NO:129)
GR-N39: CACGAGTGCTCGGAGGAGATCTAATTTTACGAGGTGTC
(SEQ ID NO:130)
GR-N40: CGGAAGAGCTGCTCAGCCGGTGAGGTCTCCCACTGGAT
(SEQ ID NO:131)
GR-N41: GCAGTGCTGGTCTGGCGCCGATGATCACCTCTTTTTCTCGGTT
(SEQ ID NO:132)
GR-N42: CATCGTGAGCTTTCTCTCTGTGAAAATTTTTGATATCCG
(SEQ ID NO:133)

RdVer5 with flanking sequence of pRAM to end of Sfi I primers
1) coding strand

RD-C1: GGAACAGGATCCCATGATGAGCCGAGGAGAAAAATGAT
(SEQ ID NO:134)
RD-C2: TATGCGGCTGACCTCCTTCCCTTCCCTTTGGAGATTTGACT
(SEQ ID NO:135)
RD-C3: GCGCGGAATACGCTTTGGTGGTGGCTCCTCGCGAACACCGTCTCC
(SEQ ID NO:136)
RD-C4: ATTTGCGCTACGCTTTGGTGGATGCTGCGGGATGATATC
(SEQ ID NO:137)
RD-C5: TTGTGAGCTCACAAAGAGGTTTTTTTTGCGGACCGTCTGTG
(SEQ ID NO:138)
RD-C6: GCTCGATGCTTCCTCCACCAATGATGCTGTTAAAGGAAAGA
(SEQ ID NO:139)
RD-C7: TCCTTATATTTGATCTGATGAAAAAATAATTGCCGTTTCCTCAT
(SEQ ID NO:140)
RD-C8: TCCAGATCGTCGCGCAGTATGATCATCGTGATGTGCT
(SEQ ID NO:141)
RD-C9: CCGTACGCTACGGAAGATGCTACATTCACCCGACGACTGAGA
(SEQ ID NO:142)
RD-C10: TCTATGATCTTCTCTAGGCACAGATTCGCTCCTGACCTCAA
(SEQ ID NO:143)
RD-C11: GAAATTCTGAAACAAAGTCTCGGAAAGTCCAAAGGCCAC
(SEQ ID NO:144)
RD-C12: AGCTTATGACCTGCTACTGAAGGTCACTGCTGGAGA
(SEQ ID NO:145)
RD-C13: ATATCTACGCGGCTGCTACCTGCTTTGCGCTATCTCTGC
(SEQ ID NO:146)
RD-C14: CTTATACGCAATGGATATCGACACTGAAATTTAAACACTCCAC
(SEQ ID NO:147)
RD-C15: TCTGACCTCCGTGAAACAGATGTCGACCCCTTCGATAGCA
(SEQ ID NO:148)
RD-C16: GCCGCTACTGATCGCTCCCAAGGAAAGTCTACGCAACCCAC
(SEQ ID NO:149)
RD-C17: TCAAACACATGTGCTGCTGCTGATGGATCAGCTGCTGATGCA
(SEQ ID NO:150)
RD-C18: GCCTACGAGCCTGCTGGATTCGGCTCGGATCGCTTGGG
(SEQ ID NO:151)
RD-C19: TCTACTGCCCTTTCTCTAGACCTGCTTGGGCTTTCATTAC
(SEQ ID NO:152)
RD-C20: TTGGGTTACCTTTCTATAGGTGTGTCCTCCGGGTATGTTCTC
(SEQ ID NO:153)
RD-C21: CGGCTTTTTCATGAGGAGGCTTTCTTGAAAAGGATCCATCACAG
(SEQ ID NO:154)
RD-C22: ATATTAAGAGTGCCGACGTTATCAACAGTGCTGACG
(SEQ ID NO:155)
RD-C23: CCTGGTTTTTTGGCTAAGAGCCCAACAGTGAGGAGAAGAC
(SEQ ID NO:156)
RD-C24: TTAGGTCTCTCTGCTGTAATGTGGCTGGTGGCGCTCCAC
(SEQ ID NO:157)
RD-C25: TGGCTAAAGAGGCGCTGTAAGGGCCGCAAAGGCTTGGAA
(SEQ ID NO:158)
RD-C26: TCTTCTGACGGGATGCTGTGCTGCTCCGGCTACGCAATC
(SEQ ID NO:159)
RD-C27: ACCAGGCTATTATCACCTGCTTACCCCGCTGTAGTGAAG
(SEQ ID NO:160)
RD-C28: GCGGCTTTTGGCCGCTGTCACCATCAGTCTCTGCTAA
(SEQ ID NO:161)
RD-C29: GATCGCTGACCGAAATGGTGAAGCTGTTTGGGCCCTAAC
(SEQ ID NO:162)
RD-C30: CAACTGGGCGACGCTGTATCAACAGCCCATATGCTGAGCA
(SEQ ID NO:163)
RD-C31: AGGTTTATGCAATACCTGCAAGATCTCAAAGAAGAGCAC
(SEQ ID NO:164)
RD-C32: CGACGAGAGACGGCTGTCATCTGGATATTTTGAAAT
(SEQ ID NO:165)
RD-C33: TACGAGAGAGAATGCAATCCTTCTACGCTGCTGATGCA
(SEQ ID NO:166)
RD-C34: AGGAGCTGATCAAATACTAAACAGGTTGACGAGTGCTTGAGCC
(SEQ ID NO:167)
RD-C35: TGAGTTGGGAGGAGATTTGTTGAAAATTTCTGCACTCCAGC
(SEQ ID NO:168)
2) non-coding strand

RD-N1: TAATCATGAGACTTTACTAACCACCGCCTTTCCACACA (SEQ ID NO:176)
RD-N2: ATGTTCTTACAGCAAGCTCTCAGGCTGTAATTTGGGCTGTAC (SEQ ID NO:177)
RD-N3: GTATGAGCGGATTAGTCAAACAAACACAGCCGCGCCACGC (SEQ ID NO:178)
RD-N4: AAGTTACTTACAGTGTCAACCTGACATGGGCACTACAT (SEQ ID NO:179)
RD-N5: ACGCTTCTTGGCGTATATTTCTTTTCACAGCGTCTTGAC (SEQ ID NO:180)
RD-N6: AAGCGAAACAGGAAAGGAGATCCGGCCGGTCGCAATTAGA (SEQ ID NO:181)
RD-N7: ATGCGGGACACACAGCCATCGGGATAGTTGGATTTATCA (SEQ ID NO:182)
RD-N8: ACAGGATCTCCACACTCACTGGAGACAACTCGTTACAC (SEQ ID NO:183)
RD-N9: CTTGATTATTATGATCAAGCCTTCTTCAAGCTCAAGCTTAA (SEQ ID NO:184)
RD-N10: AAATGTCAATCTTCTGCTTGATATACACAAAAATACCGAAT (SEQ ID NO:185)
RD-N11: GCAACACGCGCGTGCTCTGCTATGTTGCCTTGCTTGTTCT (SEQ ID NO:186)
RD-N12: GAGCTTTATGACATACCCGCTCTCGACATAGGCTGCTTG (SEQ ID NO:187)
RD-N13: ATACAGACGCTCTGCCACCTGTTGCTAAGGCCAAAGCTTAC (SEQ ID NO:188)
RD-N14: CAGTTTCTGCGGCTGCTGAGATCTTACAGAAGCTGCTAG (SEQ ID NO:189)
RD-N15: GACAGCGGCTCAAGGACCCACTCTACATCAGCGGAGA (SEQ ID NO:190)
RD-N16: GACCTGAAATATTACAGCTCGTGATATCTGCGGCGCAGA (SEQ ID NO:191)
RD-N17: AGCCCAACAGAATCCGTGAGATCTCAAGCGCTTTGGCGGCCAC (SEQ ID NO:192)
RD-N18: TTACGCGCATCCTCTGACAGCTGCTGACGCTGCGCAACAC (SEQ ID NO:193)
RD-N19: ATTACGCGCATGGAAGCACAGTGCATGCTGCAAGCTGTG (SEQ ID NO:194)
RD-N20: GCCTCTTCTAGCAAAACAGATCAGTCGATGCGATTGAT (SEQ ID NO:195)
RD-N21: GACACTGCGGCTTCATACACTTGGAGATGGCTGATTTCCAA (SEQ ID NO:196)
RD-N22: GCCTCTTCTGATCACAACAAACGCGAGCAATATACACCGGAGAC (SEQ ID NO:197)
RD-N23: CGAACATAAAATGTTACAAACAAATGATGATTGTCGAAAC (SEQ ID NO:198)
RD-N24: ATGGAAGAAGACGCAAGCTCTCTCTATCAGCGACAGA (SEQ ID NO:199)
RD-N25: ATACAGACTGCGGCGATGCGGATGAGCTGAGCTGAGCTGA (SEQ ID NO:200)
RD-N26: GACAGCGGCTCAAGGACCCACTCTACATCAGCGGAGA (SEQ ID NO:201)
RD-N27: TGGGGAGTCCCACTACAGTCGCTACAGAAATGCGTAC (SEQ ID NO:202)
RD-N28: ACTTTGTTCCACAGGCTGAGTTGCGGCGCAATTGAGGTT (SEQ ID NO:203)
RD-N29: CGATTTGCGCTCTGACTGGAAGAGGAGTAAAAATTGTGCCAA (SEQ ID NO:204)
RD-N30: AGATTGCGAAACCTGAAATTTTCGAGCTGACGCTGCGCAAGA (SEQ ID NO:205)
RD-N31: ATGATAGCCATTAAATAAGCTTGGCGCTTGGACTCTCCA (SEQ ID NO:206)
RD-N32: GGACTCTTCTGCAAGTTACCTGGAGTGGAGCAACAGTACCTC (SEQ ID NO:207)
RD-N33: GCCTCTTCTGAGTACCTGGAGTGGAGCAACAGTACCTC (SEQ ID NO:208)
RD-N34: ATTTATGCGCGTCTAGTGGAGCAACAGTACCTC (SEQ ID NO:209)
RD-N35: ACCATCGGGGAGTTGGAAGAAGGCGGAGTTGTATTTT (SEQ ID NO:210)
RD-N36: TCCAGCAAGAATCTACGAGCTCTCTCTATCAGCGACAGA (SEQ ID NO:211)
RD-N37: CAAATTGTTGGGAGACTGCGGACAGCAAGGCTGCTGCTGCA (SEQ ID NO:212)
RD-N38: AAAACTCTCTTGGAGTCAAGGATGACGCTGCTGCAAGAG (SEQ ID NO:213)
RD-N39: AGCAAGTCCGGAAGGCAAGGCTGCTGCTGCAAGAG (SEQ ID NO:214)
RD-N40: CGAAAGACCGGATTTCCCGGGGCACTCAAAATCCTCAACAGAT (SEQ ID NO:215)
RD-N41: GAGAGGCGCTCCGCGCAAGGAGTTGAGTTTCTCAGGCTT (SEQ ID NO:216)
RD-N42: CATCATGCGGTCCTCTGCTGCTGCTGCTGCAAGAG (SEQ ID NO:217)
Figure 7 (Cont.)

| RELLUC. SEQ | AAAATTGGAATAATATCATCAAATCGTTCTGCAGCTCT920 |
| RLUCVER1. SEQ | GATAATTGCAAGTACATCACAGCTTCGTAAGGCTCT920 |
| RLUCVER2. SEQ | AATGGGTAGTCATCAAGACCTTCGTAAGGCTCT920 |
| RLUCFINL. SEQ | AATGGGTAGTAAGTCATCAAGACCTTCGTAAGGCTCT920 |
| RELLUC. SEQ | CGAATATGAACTAAAA933 |
| RLUCVER1. SEQ | AGAAGCGAGCAGA933 |
| RLUCVER2. SEQ | AGAAGCGAGCAGA933 |
| RLUCFINL. SEQ | AGAAGCGAGCAGA933 |
## Figure 9A
Codon usage in RELLUC
(Renilla reniformis; Genbank ACCESSION:M63501; Medline:91239583)

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>11</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>5</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>8</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>4</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>8</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>1</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>12</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>6</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>3</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>9</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>12</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>2</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>6</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>3</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>5</td>
</tr>
<tr>
<td>TCC</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>6</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>4</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>5</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>11</td>
</tr>
<tr>
<td>CGG</td>
<td>Pro</td>
<td>2</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>4</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACN</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>2</td>
</tr>
<tr>
<td>AAA</td>
<td>Lys</td>
<td>21</td>
</tr>
<tr>
<td>AAG</td>
<td>Lys</td>
<td>6</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>5</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>8</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>16</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>1</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>25</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>5</td>
</tr>
<tr>
<td>GGT</td>
<td>Gly</td>
<td>10</td>
</tr>
<tr>
<td>GGC</td>
<td>Gly</td>
<td>4</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>3</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>3</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>0</td>
</tr>
<tr>
<td>TGA</td>
<td>Cys</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Trp</td>
<td>8</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>9</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>1</td>
</tr>
<tr>
<td>CAA</td>
<td>Gln</td>
<td>6</td>
</tr>
<tr>
<td>CAG</td>
<td>Gln</td>
<td>1</td>
</tr>
<tr>
<td>CTT</td>
<td>Trp</td>
<td>4</td>
</tr>
<tr>
<td>CGC</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>2</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>2</td>
</tr>
<tr>
<td>AGT</td>
<td>Ser</td>
<td>2</td>
</tr>
<tr>
<td>AGC</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>2</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>3</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>16</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>1</td>
</tr>
<tr>
<td>GAA</td>
<td>Glu</td>
<td>25</td>
</tr>
<tr>
<td>GAG</td>
<td>Glu</td>
<td>5</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>0</td>
</tr>
</tbody>
</table>

**Note:** The codon usage table summarizes the frequency of each amino acid based on the codon usage in the RELLUC gene of Renilla reniformis.
### Figure 9B
Codon Usage in Rluc-final

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>Phe</td>
<td>4</td>
</tr>
<tr>
<td>TTC</td>
<td>Phe</td>
<td>12</td>
</tr>
<tr>
<td>TTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTT</td>
<td>Leu</td>
<td>3</td>
</tr>
<tr>
<td>CTC</td>
<td>Leu</td>
<td>6</td>
</tr>
<tr>
<td>CTA</td>
<td>Leu</td>
<td>0</td>
</tr>
<tr>
<td>CTG</td>
<td>Leu</td>
<td>13</td>
</tr>
<tr>
<td>ATT</td>
<td>Ile</td>
<td>3</td>
</tr>
<tr>
<td>ATC</td>
<td>Ile</td>
<td>18</td>
</tr>
<tr>
<td>ATA</td>
<td>Ile</td>
<td>0</td>
</tr>
<tr>
<td>ATG</td>
<td>Met</td>
<td>9</td>
</tr>
<tr>
<td>GTT</td>
<td>Val</td>
<td>2</td>
</tr>
<tr>
<td>GTC</td>
<td>Val</td>
<td>8</td>
</tr>
<tr>
<td>GTA</td>
<td>Val</td>
<td>0</td>
</tr>
<tr>
<td>GTG</td>
<td>Val</td>
<td>13</td>
</tr>
<tr>
<td>TCT</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>TTC</td>
<td>Ser</td>
<td>10</td>
</tr>
<tr>
<td>TCA</td>
<td>Ser</td>
<td>1</td>
</tr>
<tr>
<td>TCG</td>
<td>Ser</td>
<td>0</td>
</tr>
<tr>
<td>CCT</td>
<td>Pro</td>
<td>11</td>
</tr>
<tr>
<td>CCC</td>
<td>Pro</td>
<td>3</td>
</tr>
<tr>
<td>CCA</td>
<td>Pro</td>
<td>4</td>
</tr>
<tr>
<td>CCG</td>
<td>Pro</td>
<td>0</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>4</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>Thr</td>
<td>9</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>11</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>9</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>0</td>
</tr>
<tr>
<td>TAT</td>
<td>Tyr</td>
<td>2</td>
</tr>
<tr>
<td>TAC</td>
<td>Tyr</td>
<td>11</td>
</tr>
<tr>
<td>TAA</td>
<td>Tyr</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>Tyr</td>
<td>0</td>
</tr>
<tr>
<td>CAT</td>
<td>His</td>
<td>2</td>
</tr>
<tr>
<td>CAC</td>
<td>His</td>
<td>8</td>
</tr>
<tr>
<td>CAA</td>
<td>His</td>
<td>3</td>
</tr>
<tr>
<td>CAG</td>
<td>His</td>
<td>4</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>2</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>11</td>
</tr>
<tr>
<td>AAA</td>
<td>Asn</td>
<td>4</td>
</tr>
<tr>
<td>AAG</td>
<td>Asn</td>
<td>23</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>6</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Asp</td>
<td>10</td>
</tr>
<tr>
<td>GAG</td>
<td>Asp</td>
<td>28</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>1</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>2</td>
</tr>
<tr>
<td>TGA</td>
<td>Cys</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Cys</td>
<td>8</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CCG</td>
<td>Arg</td>
<td>7</td>
</tr>
<tr>
<td>CCA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>3</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>7</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>2</td>
</tr>
<tr>
<td>AGC</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>7</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>TGA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Asp</td>
<td>8</td>
</tr>
<tr>
<td>TAA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>Asp</td>
<td>1</td>
</tr>
<tr>
<td>CAT</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>CAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Asn</td>
<td>2</td>
</tr>
<tr>
<td>CAG</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>AAA</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAG</td>
<td>Asn</td>
<td>4</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>6</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Asp</td>
<td>10</td>
</tr>
<tr>
<td>GAG</td>
<td>Asp</td>
<td>28</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>1</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>2</td>
</tr>
<tr>
<td>TGA</td>
<td>Cys</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Cys</td>
<td>8</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CCG</td>
<td>Arg</td>
<td>7</td>
</tr>
<tr>
<td>CCA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>3</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>7</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>2</td>
</tr>
<tr>
<td>AGC</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>7</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>TGA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Asp</td>
<td>8</td>
</tr>
<tr>
<td>TAA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>Asp</td>
<td>1</td>
</tr>
<tr>
<td>CAT</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>CAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Asn</td>
<td>2</td>
</tr>
<tr>
<td>CAG</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>AAA</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAG</td>
<td>Asn</td>
<td>4</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>6</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Asp</td>
<td>10</td>
</tr>
<tr>
<td>GAG</td>
<td>Asp</td>
<td>28</td>
</tr>
<tr>
<td>TGT</td>
<td>Cys</td>
<td>1</td>
</tr>
<tr>
<td>TGC</td>
<td>Cys</td>
<td>2</td>
</tr>
<tr>
<td>TGA</td>
<td>Cys</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Cys</td>
<td>8</td>
</tr>
<tr>
<td>CGT</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CCG</td>
<td>Arg</td>
<td>7</td>
</tr>
<tr>
<td>CCA</td>
<td>Arg</td>
<td>0</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>3</td>
</tr>
<tr>
<td>ACT</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>ACC</td>
<td>Thr</td>
<td>7</td>
</tr>
<tr>
<td>ACA</td>
<td>Thr</td>
<td>2</td>
</tr>
<tr>
<td>AGC</td>
<td>Thr</td>
<td>1</td>
</tr>
<tr>
<td>GCT</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCC</td>
<td>Ala</td>
<td>7</td>
</tr>
<tr>
<td>GCA</td>
<td>Ala</td>
<td>3</td>
</tr>
<tr>
<td>GCG</td>
<td>Ala</td>
<td>4</td>
</tr>
<tr>
<td>TGA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TGG</td>
<td>Asp</td>
<td>8</td>
</tr>
<tr>
<td>TAA</td>
<td>Asp</td>
<td>0</td>
</tr>
<tr>
<td>TAG</td>
<td>Asp</td>
<td>1</td>
</tr>
<tr>
<td>CAT</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>CAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>CAA</td>
<td>Asn</td>
<td>2</td>
</tr>
<tr>
<td>CAG</td>
<td>Asn</td>
<td>1</td>
</tr>
<tr>
<td>AAT</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAC</td>
<td>Asn</td>
<td>7</td>
</tr>
<tr>
<td>AAA</td>
<td>Asn</td>
<td>3</td>
</tr>
<tr>
<td>AAG</td>
<td>Asn</td>
<td>4</td>
</tr>
<tr>
<td>GAT</td>
<td>Asp</td>
<td>6</td>
</tr>
<tr>
<td>GAC</td>
<td>Asp</td>
<td>11</td>
</tr>
<tr>
<td>GAA</td>
<td>Asp</td>
<td>10</td>
</tr>
<tr>
<td>GAG</td>
<td>Asp</td>
<td>28</td>
</tr>
</tbody>
</table>
**Figure 10**

Oligonucleotides for the assembly of synthetic *Renilla* luciferase gene

<table>
<thead>
<tr>
<th>Sense Strand</th>
<th>Oligo name</th>
<th>Oligo sequence from 5' to 3'</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLS1 (1-40)</td>
<td>ACCATGCGCCCAAGCTGTCATACGCACCCAGATCCGAAAA (SEQ ID NO:246)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS2 (41-80)</td>
<td>CGATGACGACCGGGCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:247)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS3 (81-120)</td>
<td>AAAAAGGCTGGGCTGACCTCTACACATGACCTGATACCT (SEQ ID NO:248)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS4 (121-170)</td>
<td>CGAGAAGGGGGGACCGTGCTGAACCTTTGCTGTAACGCGGCTCCTG (SEQ ID NO:249)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS5 (171-210)</td>
<td>CCAGCTCCTCGGTGACCGCAGCTGTCCTGCTCACTCGAGCC (SEQ ID NO:230)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS6 (211-250)</td>
<td>CTGCGCCCATGCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:251)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS7 (251-290)</td>
<td>AAGTCTGGCCGAAGGCCGGAATGTCCTACGCACTCGGGCTCCTG (SEQ ID NO:252)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS8 (291-330)</td>
<td>ATACATCAAGGTGACCGCAGCTGTCCTGCTCACTCGAGCC (SEQ ID NO:253)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS9 (331-370)</td>
<td>CCTCGGAAGAAAATTGTCTTGGTGCGGACCGAGCTG (SEQ ID NO:254)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS10 (371-410)</td>
<td>GGGTTCGGGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:255)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS11 (411-450)</td>
<td>AGATCGGACGAGGCGTGTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:256)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS12 (451-490)</td>
<td>GCTCCGACCGGCGTGTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:257)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS13 (491-533)</td>
<td>CCTGTCAGTACGACGGCGGCGTGTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:258)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS14 (534-575)</td>
<td>AAATACGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:259)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS15 (576-615)</td>
<td>TGCGGAAACGAGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:260)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS16 (616-655)</td>
<td>TACCGAGGAGGGCGGGAATGTCCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:261)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS17 (656-700)</td>
<td>GCCGCTGGGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:262)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS18 (701-740)</td>
<td>GTTCGGGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:263)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS19 (741-780)</td>
<td>CGCCGCGGAGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:264)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS20 (781-820)</td>
<td>TTGGGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS21 (821-860)</td>
<td>TCCCTAAGACGAGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:266)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS22 (861-900)</td>
<td>TCACCGGAGGAGGGCGGGAATGTCCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:267)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLS23 (901-949)</td>
<td>CACAGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:268)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Anti-sense Strand**

<table>
<thead>
<tr>
<th>Oligo name</th>
<th>Oligo Sequence from 5' to 3'</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLSA1 (1-29)</td>
<td>GCACATGAGGAAATGGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:269)</td>
<td></td>
</tr>
<tr>
<td>RLSA2 (30-69)</td>
<td>CGGCGTGGCTGGATGACCGCAGCTGTCCTGCTCACTCGAGCC (SEQ ID NO:270)</td>
<td></td>
</tr>
<tr>
<td>RLSA3 (70-109)</td>
<td>TGCGGACGAGGGCGGGAATGTCCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:271)</td>
<td></td>
</tr>
<tr>
<td>RLSA4 (110-149)</td>
<td>ACGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:272)</td>
<td></td>
</tr>
<tr>
<td>RLSA5 (150-189)</td>
<td>TAGTCTGGGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:273)</td>
<td></td>
</tr>
<tr>
<td>RLSA6 (190-229)</td>
<td>CTTAGGAGCAAGGAGCTTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:274)</td>
<td></td>
</tr>
<tr>
<td>RLSA7 (230-269)</td>
<td>TTGGGGCAGAATATGGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:275)</td>
<td></td>
</tr>
<tr>
<td>RLSA8 (270-309)</td>
<td>CGGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:276)</td>
<td></td>
</tr>
<tr>
<td>RLSA9 (310-349)</td>
<td>ACGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:277)</td>
<td></td>
</tr>
<tr>
<td>RLSA10 (350-394)</td>
<td>ACGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:278)</td>
<td></td>
</tr>
<tr>
<td>RLSA11 (395-434)</td>
<td>GTTCCGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:279)</td>
<td></td>
</tr>
<tr>
<td>RLSA12 (435-474)</td>
<td>CTTCCGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:280)</td>
<td></td>
</tr>
<tr>
<td>RLSA13 (475-517)</td>
<td>AGCGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:281)</td>
<td></td>
</tr>
<tr>
<td>RLSA14 (518-559)</td>
<td>CGTTGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:282)</td>
<td></td>
</tr>
<tr>
<td>RLSA15 (560-599)</td>
<td>TAGTTGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:283)</td>
<td></td>
</tr>
<tr>
<td>RLSA16 (600-639)</td>
<td>AGGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:284)</td>
<td></td>
</tr>
<tr>
<td>RLSA17 (640-679)</td>
<td>GTGTTGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:285)</td>
<td></td>
</tr>
<tr>
<td>RLSA18 (680-719)</td>
<td>TCCCCGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:286)</td>
<td></td>
</tr>
<tr>
<td>RLSA19 (720-754)</td>
<td>GGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:287)</td>
<td></td>
</tr>
<tr>
<td>RLSA20 (765-804)</td>
<td>TTCCGGAGAATAGTACATCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:288)</td>
<td></td>
</tr>
<tr>
<td>RLSA21 (805-849)</td>
<td>CAGGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:289)</td>
<td></td>
</tr>
<tr>
<td>RLSA22 (850-889)</td>
<td>GGGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:290)</td>
<td></td>
</tr>
<tr>
<td>RLSA23 (890-929)</td>
<td>TGGGAGGATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:291)</td>
<td></td>
</tr>
<tr>
<td>RLSA24 (930-969)</td>
<td>ACATCGGCTTTCGAGACTGACCTACGCACTCAGGTGCTGCAAAGC (SEQ ID NO:292)</td>
<td></td>
</tr>
</tbody>
</table>
Renilla luciferase gene in pGL3 series

SV40 promoter   RL gene   SV40 enhancer

Control        Basic        Enhancer        Promoter

Fig BA
Figure 13: RL Co-Reporter Vector Series

- pRL-TK
  - HSV TK promoter
  - Intron
  - T7 promoter

- pRL-CMV
  - CMV Immediate Early Enhancer/Promoter

- pRL-SV40
  - SV40 Early Enhancer/Promoter

- pRL-null
  - MCS
Figure 13 (Continued)
Half-life of RL-synthetic and RL-native in CHO Cells

\[
y = -0.0872x + 1.9693 \\
R^2 = 0.9621
\]

\[
y = -0.0974x + 1.9659 \\
R^2 = 0.9739
\]

Time after addition of Cycloheximide (hour)

Ig(percentage of luciferase activity)
RLU

Reaction time (minutes)

TNT (RL-final versus RL-native, linear range)

-500000
0
500000
1000000
1500000
2000000

R² = 0.9938

RL-native

y = 119195x - 1E+06

R² = 0.9938

RLU 500000 -500000

y = 30467x - 596247

R² = 0.9938

RL-final

In vitro translation of RNAs of native RL and RL-final (30°C)
In vitro translation of RNAs of native RL and RL-final (30°C, linear range)

- RL-native
- RL-final

\[
y = 4426x - 17713 \quad R^2 = 0.9969
\]

\[
y = 2001.2x - 9782.2 \quad R^2 = 0.9902
\]
In vitro translation using wheat germ extract

Figure E
In vitro translation using wheat germ extract

\[ R^2 = 0.9974 \]
\[ y = 89.467x - 21.129 \]

\[ R^2 = 0.9986 \]
\[ y = 458.67x - 11393 \]
Effect of firefly expression with increasing amounts of TK vector co-transfected

Effect of firefly expression with increasing amounts of TK vector co-transfected
Beetle Luciferin + ATP + O$_2$ $\xrightarrow{\text{Firefly Luciferase}}$ Oxy Luciferin

\begin{align*}
\text{Beetle Luciferin} & \quad \text{Firefly Luciferase} \\
\text{Oxy Luciferin} & + \text{AMP + PP} + \text{CO}_2 + \text{Light}
\end{align*}
GRver5.1 DNA sequence of pGL3 vectors

ATGGTGAAACGCAGAAAAGAAGCTGTATCTACGCCCAAGACACCACCTGACCTCC 50
ACTGAAGAGCTCACTCGGTGATGAGTTCCTTCGAGACACTGCACTGCTGCTGGCTGGACCTAAGCCTGCC 100
ATATGCACTCCCTCTCTCAAGAGACTCTGTCGACTTGCTGGTGAGGAGCCAGAAGCTGCTGCTGCTGCACTGCTG 150
TCCTCAAAAGATTTTCTGACCTACTTCACCTGATGCTGTTGGGCCCAGACTCCCG 200
TTGATGTTTGGTACAAGATGATGGTGGAGCATTTTGGTGGGCTGAGAATA 250
ACACTCTGTTCTTTATTCCTGTAATCGTGCTCTTGGTGACATCCGCTGATGATT 300
GTGCCCTCTGGAATGAACTTTACACTCCCAGATGACGCTGTGTAAGGTT 350
GGGTATTGAAACACTTCAAAATCTGCTTTTACTACCAAAAGCATTGACATT 400
AGGTCTTGGGAAATCCCAAGTTCTCTGCTATATCTTCAACCAGCCTATT 450
CTGGAATCCGTCGAAAACATCCACCGCTGCGAGAGCCTCCCTAACTCTAT 500
CTCTCGTTACAGCTGATTAAATATCGCTAATTTCAAGCCCTTGTCAATT 550
ATCCAGTCGAGCAAGCTGCGCTGATTTTGTGCCTCCGGCAACCAGCTG 600
TGACCTAAAGGTGATGGTAGACACTGACCAAGAATATCTGGTGTGGTGTCAG 650
CCACGCTCTGCACTCTCTGTTGTACCTCAATTGATACCCTTGTGGTGACTG 700
TGCTGTTGTATCTGCGCTTTTCTTCACGCCCCCTTGGTTTCCTGATTACCTG 750
GCCATATTCTCTCTGCTGCCGCTGGTGCTGCTACATGTTCCTGTCTCGTGACCA 800
AGAAGCTTTCTGGAAAGCTATTCTCACAGACTCGAGGCTGCTCGTGATCA 850
AGCTCCCTCTCTGCACTTCATTTCTGGGTCAAGAATCCTCTCTTTTGGTGACAG 900
TATGTATCTGTGAGCAGCTTGGTGGTGACGGTGTGGTGCTGCTGCTCGCTCTTTTGG 950
CAGAGAGATCTGGCGAGTGATCTGGCTGTCGACGTGCGTTCCCTGGATACCT 1000
GCTGCGCTTCTGTTTGAGGACTGAAAGACCTCTTCTGCTAACACTCCATAGCT 1050
CGAGACAGTTTTGAAGCTGATTACCTTGGCTGGTGACCTCTCTGTATGCG 1100
TGCCAAGATGGCGAGCCGAGCTGACGGGCAAGACACTGGGCCCCAGATCAC 1150
TCGATGAAATGTGTTATTAAGGGCCCTATGTTCTCTAAAGGCTACGTGAAAC 1200
AAGTGATGAGGCGCTACTAAGAGGCCATTGATGATGATGCTGGCTGCTCCATAAG 1250
CGCGACCTCTGCTTTACTATGTATAAGGAGGCCAGAAGACTCTCTATGTGTCTGACTC 1300
GCTACAAGAATTTGATATTAAGGCTTAAAGGCTTGACCACAGCAGA 1350
CTGGAAGAATTTTGCTGAGAAGACCTCTTGATGCTTCCGGCACTGGCGCTGCTGTG 1400
GGGTATTGGAGCTGTTGCCGGATTGGCATTGCCCCATTTTGCCGCTGAGA 1450
AACACCGGCAAGGAGTACTGCTGTAAGGAGGAGCTACAGACTATATGCGCC 1500
GAGGGCTGTCTCAGACCCAAATATCTGGTGGGGGCGGCTCTGCGCTGCTG 1550
TTCTATTTCCAGCGAGGTACCCTGGTAAGATACCTGCTAAAGAGTGGCTGA 1600
AGCAACTCTCAGAAAAGAGCTGCGGCC 1626
RDver5.1 DNA sequence of pGL3 vectors

ATGGTGAAGCCTGAGA...AAATGTATCTATATGGCCCTGAGCCTCTCCATCC
TTTGGGAGAATTTGACCTCGGCGGAAATGTATGTCTCTCGTGCTCGAAGCC
ACTCTcATTTTGGCTCGATGGATGTTGTCGCCGAATATCTCATCC
AGCTCAAAGGAA...TTTGGAGAGGCAACCGTCTTGGCTCGATCCATCC
CAAAGTGGCTCACAAGAATGGAACGCGTCTTGGCTCGATCCATCC
ATACCGGTTCTTCTCATACGACTCCGCGATCGGTTATGCTGAT
GTGGCTCCAGACGCAAGGCTTCCAGATCCAGTCCAGGAATGTGTAAGC
GGTATCTCTAAAGCCAC...TTGGCTCTCCATACGACTCCGCGATCCATCC
ACCCTGTTGAACAAAGTTGACCCATTGCTTGACGATCGCTACTCTGGA
CTCCCAAGGGAGCTATGCGAAGCCATCAAACATTTTGCGTGCCTGAT
CCGAGTCTCGATCCAGCCTTTGGGGCTCTGTCGATCCATCC
TTCTGGCTCTTTGGCTCTTCATAGGTATATTACCTTTGGCGGATTTTGATCA
GGAGGCTCTTTGGGAAGGGCATCAAAGATTTAGGAGCGTCTGCTGTCAC
ACGAGCTCATTGGCATGATTTTGCTTTGGGTACGAGCCACGAGGGATCAG
TACGACTTGTCTCTACTCAGCTGGAATTTGCTCTGGCTCCTCAGTGC
TAAGGAGTCTCGAGAAGTGCGCGCCCAAGCCTGGTAATCTCCAGGATTC
GATGTTGCTCTCGGCTCCAGCAATCTACGCGCTATATTATCGCTCCTC
CGGCGATTACCGGCTCTTTGGGGCGTGTATCTCCAGTCTGCGG
TGCTAAGATCGTCTGATCGGAACCTGGTAAAGGCTTTGGGCCGGAACCAAG
TGGCGAGCTGTATCAAGGCCTATGTGAGCAAGGTTGTAATGCTCAAT
AAGCTTGAGACCGGAGCCATCGAGACGCCGCTGTTGCAATCC
TGGTGAATTGGATATTACGAGAGATGGACATTATTACGCTGTGGATC
GTACAAGGAGCTGATCAAAATACAGGTTAGCCGAGGCTGCTCAGCTGA
TGGGAGGAGATTTGCTGGAAATACGATAGGTACATTGCCTGGTGTG
CGGCTATTCTCGATCGAGCGCGGCCAACGCGGCCCTCTGCTTTGGGGTC
AGCAGCTGTGGAAGAATATTACGGCGCAAGGAGATGTATGATGATTACCTGGCT
GAAGTGTTGCGCATATCAACTGACTGTTGGCGCGTCGGTTGGGTGGTTGGA
CTCCATCCCTCGTAAACGTGACGGCAAATAATTACCGCGCAAGGAAGCTG
AACATTTGTTGGAGAAAGCCCGCGGCTT

SEQ ID NO: 294

1626
RD1581H9 DNA sequence of pGL3 vectors

ATGGTAAAGCGTGAAGAAAATGCTATCTATGCGCCCTGAGCCTTCCATCC  50
TTTGGAGGATTGACTGCCGCGCAATACTGTTTGCCTGCTCTCCGAACGC  100
ACTCTCATTTTGGCTCAAGCCTTGGTGGATGGTGCGGAGATATCTTTG  150
AGCTACAAGGAGTTTTTTGAGGCAACCTGTTGCTGAGCTCCTCCCA  200
CAAATGGTGCTACAAAGATGAAACAGACATGGTTAATGATCTGCTGGAAACA  250
ATACCCGTCTCTCTATCCAGCCGACATGGTAGTTGCGGATATCTTG  300
GGTGGCTCCAGCTCAAGGAGCTACATTCCGCAACGACTGTTAAAGTCAT  350
GGGATATCATGACAGCATAGTGGTCTTCACCAACTCACAAGAATATCTGGAACA  400
AACCTTGGAAGAACACGACATATGTCTGTTAGCTACAGGTAATC  450
CTCCCAAAAGGAGCTATGCAAGACCCCATATAAAACATTTGCCGCTGTGAT  500
CTCTGCTATTCCAGCCAGCCACACATGCGAAACTTTAAAACACTCACCCTCG  550
ACCCCTGGAACAAGCGGCGATCTCTGTTAGCTACAGGTAATC  600
CTCCCAAAAGGAGCTATGCAAGACCCCATATAAAACATTTGCCGCTGTGAT  650
CCATGCTCTGATCCACAGCCTACGCTACGCAACTGCTTCTGGTGCTACGC  700
TCCTGGCTCTACTGCTGCTTTCTCCATGTGCCTTGGCTCTTCTCTCTG  750
GCTATTTATGGTGCTGCTCCGCGCTGATTAGTTGCTCGCGCTTTTGTACCA  800
GAGGGCTTTCTGAAACCCATCAAGATTAGTGTAGCTGAGCGTCATCA  850
ACGTGCTAGCGTKATCCTGGTTTGGGTACACGACCTCGTGAGACAAG  900
TACGACTTGTCTCTCACAGTCGTAATGTGTTGGCTGGTGCGCTCCACCTCG  950
TAAAGGAGTGTGCTGAAAGTGCCCGCAAAAGCTGTGAAATCTTCCAGGATTC 1000
GTGTTGCTTTGGCTCTTACACCAGATCGGATTACCCAGGATCTCCT 1050
GGGATGATTTTAAGGAGCGGCCTTTGGGCGGTGTCATCCACACTATGC 1100
TGCTAAGATGCTGTAGTCCGGAACACTGTGTTGCGGCCGACCAAG 1150
TGCGGGGCTGCTGTATCAAAAGTCGTTGAAGCTTGTTGGCGCAACCAAG 1200
AACAGTTGGAAGCATACCAAAAGGAGCGACATGCGACAGCGGCTGGATGCA 1250
TGGGTATTTTTGGAATACAGGAGATAGAATAGGATATTATACGTCGTTGGATC 1300
GTTCAGAAAGCGTCTGACCAATACAAAGGTAGACGAGGCTGTGCTCCACAGTGA 1350
TTGGAGGAGATTTGCTTGGAAAATAGCTGATCCCTGCGCGATGTGCTGAT 1400
CGGCATCTGCTGATCTGGAGCCCGCGCAACTGCTCTTCTGTTGGTGTCA 1450
AGCAAGGCTGACAAAGATAACCCGCCAAAGAAGAGTAATGATATACCCTGGCT 1500
GAAGTGTGAGACAGCATATATATTTACATGGCTGCGCGCTTGCTGTGCTGAT 1550
CTCCCATCCTCGTAACGTAACAGGCAAAATACCAGGGCAAGGAGCTGTGTTGA 1600
AAACATTTGGGTGTAAGGCGGCGCGGCT  1626
## GRver5.1 protein sequence of pGL3 vectors

| MVKREKNVGYPEPLHLEDLTAGMLFRALRKHSLHQLPVQVPGDESL | 50 |
| SYKEFFPSEATVLLASLHNCGGYGDNVDVSCANENNTTFIPVIAAWYQMA | 100 |
| VAPVESYIDPCKVNLGISKQPIVMTNKILNVLEQSGKTRFKNKRII | 150 |
| LDTVENHCESLPNFSRYSDNSGANFFPKLFDPVEQAIAISSLGGTG | 200 |
| LPGVQWGQTQMICVRLHALDPRVGTQVLIPVLVPYLPVFHPFSITL | 250 |
| GYFMVGLRVMFTFQDFAFLEIKAQDIYDEYRVSINVPVSLFSLKGPLVDK | 300 |
| YLDSLLRSLCLCGAALPAKEVAAVEAAKLNLPGIRCPGQLTESTSSNTHSL | 350 |
| RDEFKSOSLSGVRTPLMMAKIAADRETQKALNPQVQGCLIKGMVPSKGYVN | 400 |
| NKVEATKEADDD DWLHSGDGFYYDEDEHPPVVDYREKLYKQSQVAPAE | 450 |
| LEEILLNKCPRIDRCVAVPVIDLEAGELPSAFVQKPQKEITAKEYDYLA | 500 |
| ERVSHTKYLHGGVRFVDSIПрНТЬКИТРКЕЛККЛЖЕКАГГ | 542 |

## RDver5.1 protein sequence of pGL3 vectors

| MVKREKNVGYPEPLHLEDLTAGMLFRALRKHSLHQLPVQVPGDESL | 50 |
| SYKEFFPSEATVLLASLHNCGGYGDNVDVSCANENNTTFIPVIAAWYQMA | 100 |
| VAPVESYIDPCKVNLGISKQPIVMTNKILNVLEQSGKTRFKNKRII | 150 |
| LDTVENHCESLPNFSRYSDNSGANFFPKLFDPVEQAIAISSLGGTG | 200 |
| LPGVQWGQTQMICVRLHALDPRVGTQVLIPVLVPYLPVFHPFSITL | 250 |
| GYFMVGLRVMFTFQDFAFLEIKAQDIYDEYRVSINVPVSLFSLKGPLVDK | 300 |
| YLDSLLRSLCLCGAALPAKEVAAVEAAKLNLPGIRCPGQLTESTSSNTHSL | 350 |
| RDEFKSOSLSGVRTPLMMAKIAADRETQKALNPQVQGCLIKGMVPSKGYVN | 400 |
| NKVEATKEADDD DWLHSGDGFYYDEDEHPPVVDYREKLYKQSQVAPAE | 450 |
| LEEILLNKCPRIDRCVAVPVIDLEAGELPSAFVQKPQKEITAKEYDYLA | 500 |
| ERVSHTKYLHGGVRFVDSIПрНТЬКИТРКЕЛККЛЖЕКАГГ | 542 |

## RD1561H9 protein sequence of pGL3 vectors

| MVKREKNVGYPEPLHLEDLTAGMLFRALRKHSLHQLPVQVPGDESL | 50 |
| SYKEFFPSEATVLLASLHNCGGYGDNVDVSCANENNTTFIPVIAAWYQMA | 100 |
| VAPVESYIDPCKVNLGISKQPIVMTNKILNVLEQSGKTRFKNKRII | 150 |
| LDTVENHCESLPNFSRYSDNSGANFFPKLFDPVEQAIAISSLGGTG | 200 |
| LPGVQWGQTQMICVRLHALDPRVGTQVLIPVLVPYLPVFHPFSITL | 250 |
| GYFMVGLRVMFTFQDFAFLEIKAQDIYDEYRVSINVPVSLFSLKGPLVDK | 300 |
| YLDSLLRSLCLCGAALPAKEVAAVEAAKLNLPGIRCPGQLTESTSSNTHSL | 350 |
| RDEFKSOSLSGVRTPLMMAKIAADRETQKALNPQVQGCLIKGMVPSKGYVN | 400 |
| NKVEATKEADDD DWLHSGDGFYYDEDEHPPVVDYREKLYKQSQVAPAE | 450 |
| LEEILLNKCPRIDRCVAVPVIDLEAGELPSAFVQKPQKEITAKEYDYLA | 500 |
| ERVSHTKYLHGGVRFVDSIПрНТЬКИТРКЕЛККЛЖЕКАГГ | 542 |
SYNTHETIC NUCLEIC ACID MOLECULE COMPOSITIONS AND METHODS OF PREPARATION

STATEMENT OF GOVERNMENT RIGHTS

[0001] The invention was made at least in part with a grant from the Government of the United States of America (grant DMI-9402762 from the National Science Foundation). The Government has certain rights to the invention.

BACKGROUND OF THE INVENTION

[0002] Transcription, the synthesis of an RNA molecule from a sequence of DNA is the first step in gene expression. Sequences which regulate DNA transcription include promoter sequences, polyadenylation signals, transcription factor binding sites and enhancer elements. A promoter is a DNA sequence capable of specific initiation of transcription and consists of three general regions. The core promoter is the sequence where the RNA polymerase and its cofactors bind to the DNA. Immediately upstream of the core promoter is the proximal promoter which contains several transcription factor binding sites that are responsible for the assembly of an activation complex that in turn recruits the polymerase complex. The distal promoter, located further upstream of the proximal promoter also contains transcription factor binding sites. Transcription termination and polyadenylation, like transcription initiation, are site specific and encoded by defined sequences. Enhancers are regulatory regions, containing multiple transcription factor binding sites, that can significantly increase the level of transcription from a responsive promoter regardless of the enhancer’s orientation and distance with respect to the promoter as long as the enhancer and promoter are located within the same DNA molecule. The amount of transcript produced from a gene may also be regulated by a post-transcriptional mechanism, the most important being RNA splicing that removes intervening sequences (introns) from a primary transcript between splice donor and splice acceptor sequences.

[0003] Natural selection is the hypothesis that genotype-environment interactions occurring at the phenotypic level lead to differential reproductive success of individuals and therefore to modification of the gene pool of a population. Some properties of nucleic acid molecules that are acted upon by natural selection include codon usage frequency, RNA secondary structure, the efficiency of intron splicing, and interactions with transcription factors or other nucleic acid binding proteins. Because of the degenerate nature of the genetic code, these properties can be optimized by natural selection without altering the corresponding amino acid sequence.

[0004] Under some conditions, it is useful to synthetically alter the natural nucleotide sequence encoding a polypeptide to better adapt the polypeptide for alternative applications. A common example is to alter the codon usage frequency of a gene when it is expressed in a foreign host cell. Although redundancy in the genetic code allows amino acids to be encoded by multiple codons, different organisms favor some codons over others. It has been found that the efficiency of protein translation in a non-native host cell can be substantially increased by adjusting the codon usage frequency but maintaining the same gene product (U.S. Pat. Nos. 5,096, 825, 5,670,356, and 5,874,304).

[0005] However, altering codon usage may, in turn, result in the unintentional introduction into a synthetic nucleic acid molecule of inappropriate transcription regulatory sequences. This may adversely effect transcription, resulting in anomalous expression of the synthetic DNA. Anomalous expression is defined as departure from normal or expected levels of expression. For example, transcription factor binding sites located downstream from a promoter have been demonstrated to effect promoter activity (Michael et al., 1990; Lamb et al., 1998; Johnson et al., 1998; Jones et al., 1997). Additionally, it is not uncommon for an enhancer element to exert activity and result in elevated levels of DNA transcription in the absence of a promoter sequence or for the presence of transcription regulatory sequences to increase the basal levels of gene expression in the absence of a promoter sequence.

[0006] Thus, what is needed is a method for making synthetic nucleic acid molecules with altered codon usage without also introducing inappropriate or unintended transcription regulatory sequences for expression in a particular host cell.

SUMMARY OF THE INVENTION

[0007] The invention provides a synthetic nucleic acid molecule comprising at least 300 nucleotides of a coding region for a polypeptide, having a codon composition differing at more than 25% of the codons from a wild type nucleic acid sequence encoding a polypeptide, and having at least 3-fold fewer, preferably at least 5-fold fewer, transcription regulatory sequences than would result if the differing codons were randomly selected. Preferably, the synthetic nucleic acid molecule encodes a polypeptide that has an amino acid sequence that is at least 85%, preferably 90%, and most preferably 95% or 99% identical to the amino acid sequence of the naturally-occurring (native or wild type) polypeptide (protein) from which it is derived. Thus, it is recognized that some specific amino acid changes may also be desirable to alter a particular phenotypic characteristic of the polypeptide encoded by the synthetic nucleic acid molecule. Preferably, the amino acid sequence identity is over at least 100 contiguous amino acid residues. In one embodiment of the invention, the codons in the synthetic nucleic acid molecule that differ preferably encode the same amino acids as the corresponding codons in the wild type nucleic acid sequence.

[0008] The transcription regulatory sequences which are reduced in the synthetic nucleic acid molecule include, but are not limited to, any combination of transcription factor binding sequences, intron splice sites, poly(A) addition sites, enhancer sequences and promoter sequences. Transcription regulatory sequences are well known in the art.

[0009] It is preferred that the synthetic nucleic acid molecule of the invention has a codon composition that differs from that of the wild type nucleic acid sequence at more than 30%, 35%, 40% or more than 45%, e.g., 50%, 55%, 60% or more of the codons. Preferred codons for use in the invention are those which are employed more frequently than at least one other codon for the same amino acid in a particular organism and, more preferably, are also not low-usage codons in that organism and are not low-usage codons in the organism used to clone or screen for the expression of the synthetic nucleic acid molecule (for example, E. coli).
Moreover, preferred codons for certain amino acids (i.e., those amino acids that have three or more codons), may include two or more codons that are employed more frequently than the other (non-preferred) codon(s). The presence of codons in the synthetic nucleic acid molecule that are employed more frequently in one organism than in another organism results in a synthetic nucleic acid molecule which, when introduced into the cells of the organism that employs those codons more frequently, is expressed in those cells at a level that is greater than the expression of the wild type or parent nucleic acid sequence in those cells. For example, the synthetic nucleic acid molecule of the invention is expressed at a level that is at least about 110%, e.g., 150%, 200%, 500% or more (1000%, 5000%, or 10000%) of that of the wild type nucleic acid sequence in a cell or cell extract under identical conditions (such as cell culture conditions, vector backbone, and the like).

[0010] In one embodiment of the invention, the codons that are different are those employed more frequently in a mammal, while in another embodiment the codons that are different are those employed more frequently in a plant. A particular type of mammal, e.g., human, may have a different set of preferred codons than another type of mammal. Likewise, a particular type of plant may have a different set of preferred codons than another type of plant. In one embodiment of the invention, the majority of the codons which differ are ones that are preferred codons in a desired host cell. Preferred codons for mammals (e.g., humans) and plants are known to the art (e.g., Wada et al., 1990). For example, preferred human codons include, but are not limited to, CGC (Arg), CTG (Leu), TCT (Ser), AGC (Ser), ACC (Thr), CCA (Pro), GTG (Val), ATC (Ile), ATT (Ile), AAG (Lys), AAC (Asn), CAG (Gln), CAC (His), GAG (Glu), GAC (Asp), TAC (Tyr), TGC (Cys) and TTC (Phe) (Wada et al., 1990). Thus, preferred “humanized” synthetic nucleic acid molecules of the invention have a codon composition which differs from a wild type nucleic acid sequence by having an increased number of the preferred human codons, e.g., CGC, CTG, TCT, AGC, ACC, CCA, CCT, GCC, GGC, GTG, ATC, ATT, AAG, AAC, CAG, CAC, GAG, GAC, TAC, TGC, TTC, or any combination thereof. For example, the synthetic nucleic acid molecule of the invention may have an increased number of CTG or TTC leucine-encoding codons, GTG or GTC valine-encoding codons, GGC or GGT glycine-encoding codons, ATC or ATT isoleucine-encoding codons, CCA or CCT proline-encoding codons, CGC or CGT arginine-encoding codons, ACC or TCT serine-encoding codons, ACC or ACT threonine-encoding codon, GCC or GCT alanine-encoding codons, or any combination thereof, relative to the wild type nucleic acid sequence. Similarly, synthetic nucleic acid molecules having an increased number of codons that are employed more frequently in plants, have a codon composition which differs from a wild type or parent nucleic acid sequence by having an increased number of the plant codons including, but not limited to, CGC (Arg), CTT (Leu), TCT (Ser), TCC (Ser), ACC (Thr), CCA (Pro), CCT (Pro), GCT (Ser), GGA (Gly), GTG (Val), ATC (Ile), ATT (Ile), AAG (Lys), AAC (Asn), CAA (Gln), CAC (His), GAG (Glu), GAC (Asp), TAC (Tyr), TGC (Cys), TTC (Phe), or any combination thereof (Murray et al., 1989). Preferred codons may differ for different types of plants (Wada et al., 1990).

[0011] The choice of codon may be influenced by many factors such as, for example, the desire to have an increased number of nucleotide substitutions or decreased number of transcription regulatory sequences. Under some circumstances (e.g., to permit removal of a transcription factor binding site) it may be desirable to replace a non-preferred codon with a codon other than a preferred codon or a codon other than the most preferred codon. Under other circumstances, for example, to prepare codon distinct versions of a synthetic nucleic acid molecule, preferred codon pairs are selected based upon the largest number of mismatched bases, as well as the criteria described above.

[0012] The presence of codons in the synthetic nucleic acid molecule that are employed more frequently in one organism than in another organism, results in a synthetic nucleic acid molecule which, when introduced into a cell of the organism that employs those codons, is expressed in that cell at a level which is greater than the level of expression of the wild type or parent nucleic acid sequence.

[0013] A synthetic nucleic acid molecule of the invention may encode a selectable marker protein or a reporter molecule. However, the invention applies to any gene and is not limited to synthetic reporter genes or synthetic selectable marker genes. In one embodiment of a synthetic nucleic acid molecule of the invention that is a reporter molecule, the synthetic nucleic acid molecule encodes a luciferase having a codon composition different than that of a wild type or parent Renilla luciferase or a beetle luciferase nucleic acid sequence. A synthetic click beetle luciferase nucleic acid molecule of the invention may optionally encode the amino acid valine at position 224 (i.e., it emits green light), or may optionally encode the amino acid histidine at position 224, histidine at position 247, isoleucine at position 346, glutamine at position 348 or combination thereof (i.e., it emits red light). Preferred synthetic luciferase nucleic acid molecules that are related to a wild type Renilla luciferase nucleic acid sequence include, but are not limited to, SEQ ID NO:21 (Rluver2) or SEQ ID NO:22 (Rlu-final). Preferred synthetic luciferase nucleic acid molecules that are related to click beetle luciferase nucleic acid sequences include, but are not limited to, SEQ ID NO:7 (GRver5), SEQ ID NO:8 (GR6), SEQ ID NO:9 (GRver5-1), SEQ ID NO:14 (RDver5), SEQ ID NO:15 (RD7), SEQ ID NO:16 (RDver5-1), SEQ ID NO:17 (RDver5-2) or SEQ ID NO:18 (RD156-119).

[0014] The invention also provides an expression cassette. The expression cassette of the invention comprises a synthetic nucleic acid molecule of the invention operatively linked to a promoter that is functional in a cell. Preferred promoters are those functional in mammalian cells and those functional in plant cells. Optionally, the expression cassette may include other sequences, e.g., restriction enzyme recognition sequences and a Kozak sequence, and be a part of a larger polynucleotide molecule such as a plasmid, cosmids, artificial chromosome or vector, e.g., a viral vector.

[0015] Also provided is a host cell comprising the synthetic nucleic acid molecule of the invention, an isolated polypeptide (e.g., a fusion polypeptide encoded by the synthetic nucleic acid molecule of the invention), and compositions and kits comprising the synthetic nucleic acid molecule of the invention or the polypeptide encoded thereby in suitable container means and, optionally, instruc-
The invention also provides a method to prepare a synthetic nucleic acid molecule comprising a plurality of transcription regulatory sequences in a parent (wild type or a synthetic) nucleic acid sequence that encodes a polypeptide having at least 100 amino acids to yield a further synthetic nucleic acid molecule having an increased number of transcription regulatory sequences relative to the parent nucleic acid sequence. Alternatively, the method comprises altering greater than 25% of the codons in a parent nucleic acid sequence which has a decreased number of codon distinct versions of a synthetic nucleic acid molecule. Thus, the invention provides a method for preparing at least two synthetic nucleic acid molecules that are codon distinct versions of a parent nucleic acid sequence that encodes a polypeptide. The method comprises altering a parent nucleic acid sequence to yield a second synthetic nucleic acid molecule having an increased number of a first plurality of codons that are employed more frequently in a selected host cell relative to the number of those codons present in the parent nucleic acid sequence. The parent nucleic acid sequence is also altered to yield a second synthetic nucleic acid molecule having an increased number of a second plurality of codons that are employed more frequently in the host cell relative to the number of those codons in the parent nucleic acid sequence, wherein the first plurality of codons is different than the second plurality of codons, and wherein the first and the second synthetic nucleic acid molecules preferably encode the same polypeptide. Optionally, the second synthetic nucleic acid molecule has a decreased number of transcription regulatory sequences relative to the parent nucleic acid sequence. Either or both synthetic molecules can then be further modified.
BRIEF DESCRIPTION OF THE FIGURES


[0023] FIG. 2. A nucleotide sequence comparison of a yellow-green (YG) click beetle luciferase nucleic acid sequence (YP81-6G01; SEQ ID NO:24) and various synthetic GR click beetle luciferase amino acid sequences (GRv1, SEQ ID NO:5; GRv2, SEQ ID NO:26; GRv3, SEQ ID NO:27; GRv4, SEQ ID NO:28; GRv5, SEQ ID NO:29; GRv6, SEQ ID NO:30; GRv5.1, SEQ ID NO:31) and various red (RD) click beetle luciferase amino acid sequences (RDv1, SEQ ID NO:32; RDv2, SEQ ID NO:33; RDv3, SEQ ID NO:34; RDv4, SEQ ID NO:218; RDv5, SEQ ID NO:219; RD7, SEQ ID NO:220; RDv5.1, SEQ ID NO:221; RDv5.2, SEQ ID NO:222; RD156-119, SEQ ID NO:223). All amino acid sequences are inferred from the corresponding nucleotide sequence. The amino acids enclosed in boxes are amino acids that differ from the amino acid present at the homologous position in SEQ ID NO:24.

[0024] FIG. 3. An amino acid sequence comparison of a YG click beetle luciferase amino acid sequence (YP81-6G01, SEQ ID NO:24) and various synthetic GR click beetle luciferase amino acid sequences (GRv1, SEQ ID NO:5; GRv2, SEQ ID NO:26; GRv3, SEQ ID NO:27; GRv4, SEQ ID NO:28; GRv5, SEQ ID NO:29; GRv6, SEQ ID NO:30; GRv5.1, SEQ ID NO:31) and various red (RD) click beetle luciferase amino acid sequences (RDv1, SEQ ID NO:32; RDv2, SEQ ID NO:33; RDv3, SEQ ID NO:34; RDv4, SEQ ID NO:218; RDv5, SEQ ID NO:219; RD7, SEQ ID NO:220; RDv5.1, SEQ ID NO:221; RDv5.2, SEQ ID NO:222; RD156-119, SEQ ID NO:223). All amino acid sequences are inferred from the corresponding nucleotide sequence. The amino acids enclosed in boxes are amino acids that differ from the amino acid present at the homologous position in SEQ ID NO:24.

[0027] FIG. 6. Oligonucleotides employed to prepare synthetic GR/rd luciferase genes (SEQ ID Nos. 35-245).

[0028] FIG. 7. A nucleotide sequence comparison of a wild type Renilla reniformis luciferase nucleic acid sequence Genbank Accession No. M63501 (RELUC, SEQ ID NO:19) and various synthetic Renilla luciferase nucleic acid sequences (Rlucver1, SEQ ID NO:20; Rlucver2, SEQ ID NO:21; Rluc-final, SEQ ID NO:22). The nucleotides enclosed in boxes are nucleotides that differ from the nucleotide present at the homologous position in SEQ ID NO:19.

[0029] FIG. 8. An amino acid sequence comparison of a wild type Renilla reniformis luciferase amino acid sequence (RELUC, SEQ ID NO:224) and various synthetic Renilla reniformis luciferase amino acid sequences (Rlucver1, SEQ ID NO:225; Rluc-ver2, SEQ ID NO:226; Rluc-final, SEQ ID NO:227). All amino acid sequences are inferred from the corresponding nucleotide sequence. The amino acids enclosed in boxes are amino acids that differ from the amino acid present at the homologous position in SEQ ID NO:224.

[0030] FIG. 9. Codon usage in wild-type (A) versus synthetic (B) Renilla luciferase genes. For codon usage in selected organisms, see, e.g., Wada et al., 1990; Sharp et al., 1988; Aota et al., 1988; and Sharp et al., 1987, and for plant codons, Murray et al. 1989.


[0032] FIG. 11. A nucleotide sequence comparison of a wild type yellow-green (YG) click beetle luciferase nucleic acid sequence (LUCPPLYG, SEQ ID NO:23) and the synthetic GR click beetle luciferase amino acid sequences (GRv5.1, SEQ ID NO:29) and the synthetic red click beetle luciferase nucleic acid sequences (RD156-119, SEQ ID NO:223). All amino acid sequences are inferred from the corresponding nucleotide sequence. The amino acids enclosed in boxes are amino acids that differ from the amino acid present at the homologous position in SEQ ID NO:23.

[0034] FIG. 13. pRL vector series. All of the vectors contain the Renilla wild type or synthetic gene as further described herein. FIG. 13A illustrates the Renilla luciferase gene in the pGL3 vectors (Promega Corp.). FIG. 13B illustrates the Renilla luciferase co-reporter vector series. pRL-TK has the herpes simplex virus (HSV) tk promoter; pRL-SV40 has the SV40 virus early enhancer/promoter; pRL-CMV has the cytomegalovirus (CMV) enhancer and immediate early promoter; pRL-null has MCS (multiple cloning sites) but no promoter or enhancer; pRL-TKm’ has HSV/tk promoter without an intron that is present in the other plasmids; pRL-GL3B has the pGL-3 Basic backbone (Promega Corp.); pRL-GL3 TK has the pGL3-Basic backbone with an HSV tk promoter.

[0035] FIG. 14. Half-life of synthetic (Rluc-final) and native Renilla luciferases in CHO cells.

[0036] FIGS. 15A-B. In vitro transcription/translation of Renilla luciferase nucleic acid sequences. A) t=0-60 minutes; B) linear range.

[0037] FIGS. 15C-D. In vitro translation of native and synthetic (Rluc-final) Renilla luciferase RNAs in a rabbit reticulocyte lysate. RNA was quantitated and the same amount was employed as in the translation reaction shown in FIGS. 15A-B. C) t=0-60 minutes; D) linear range.

[0038] FIGS. 15E-F. Translation of native and synthetic (Rluc-final) Renilla RNAs in a wheat germ extract. E) t=0-60 minutes; F) linear range.
[0039] FIG. 16. High expression from a synthetic Renilla nucleic acid sequence reduces the risk of promoter interference in a co-transfection assay. CHO cells were co-transfected with a constant amount (50 ng) of firefly luciferase expression vector (pGL3 control vector, with SV40 promoter and enhancer; Luc+) and a pRL vector having a native (0 ng, 50 ng, 100 ng, 500 ng, 1 μg or 2 μg) or synthetic (0 ng, 5 ng, 10 ng, 50 ng, 100 ng or 200 ng) Renilla luciferase gene.

[0040] FIGS. 17A-B. Illustrates the reactions catalyzed by firefly and click beetle (17A), and Renilla (17B) luciferases.

[0041] FIG. 18. Nucleotide and inferred amino acid sequence of click beetle luciferases in pGL3 vectors (GRLrev5.1 in pGL3, SEQ ID NO:297 encoding SEQ ID NO:298; RDelver5.1 in pGL3, SEQ ID NO:299 encoding SEQ ID NO:300; and RD156-1H9 in pGL3, SEQ ID NO:301 encoding SEQ ID NO:302). To clone GRLrev5.1, RD156.1, and RD156-1H9 nucleic acid sequences into pGL3 vectors, an oligonucleotide having an Nco I site at the initiation codon was employed, which resulted in an amino acid substitution at position 2 to valine.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0042] The term “gene” as used herein, refers to a DNA sequence that comprises coding sequences necessary for the production of a polypeptide or protein precursor.

[0043] The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence, as long as the desired protein activity is retained.

[0044] A “nucleic acid”, as used herein, is a covalently linked sequence of nucleotides in which the 3' position of the pentose of one nucleotide is joined by a phosphodiester group to the 5' position of the pentose of the next, and in which the nucleotide residues (bases) are linked in specific sequence, i.e., a linear order of nucleotides. A “polynucleotide”, as used herein, is a nucleic acid containing a sequence that is greater than about 100 nucleotides in length. An “oligonucleotide”, as used herein, is a short polynucleotide or a portion of a polynucleotide. An oligonucleotide typically contains a sequence of about two to about one hundred bases. The word “oligo” is sometimes used in place of the word “oligonucleotide”.

[0045] Nucleic acid molecules are said to have a “5'-terminus” (5' end) and a “3'-terminus” (3' end) because nucleic acid phosphorylase linkages occur to the 5' carbon and 3' carbon of the pentose ring of the substituent mononucleotides. The end of a polynucleotide at which a new linkage would be to a 5' carbon is its 5' terminal nucleotide. The end of a polynucleotide at which a new linkage would be to a 3' carbon is its 3' terminal nucleotide. A terminal nucleotide, as used herein, is the nucleotide at the end position of the 3'- or 5'-terminus.

[0046] DNA molecules are said to have “5' ends” and “3' ends” because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotides referred to as the “5' end” if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the “3' end” if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring.

[0047] As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide or polynucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being “upstream” or 5' of the “downstream” or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. Typically, promoter and enhancer elements that direct transcription of a linked gene are generally located 5' upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

[0048] The term “codon” as used herein, is a basic genetic coding unit, consisting of a sequence of three nucleotides that specify a particular amino acid to be incorporated into a polypeptide chain, or a start or stop signal. FIG. 1 contains a codon table. The term “coding region” when used in reference to structural gene refers to the nucleotide sequences that encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. Typically, the coding region is bounded on the 5' side by the nucleotide triplet “ATG” which encodes the initiator methionine and on the 3' side by a stop codon (e.g., TAG, TGA). In some cases the coding region is also known to initiate by a nucleotide triplet “TTG”.

[0049] By “protein” and “polypeptide” is meant any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation). The synthetic genes of the invention may also encode a variant of a naturally-occurring protein or polypeptide fragment thereof. Preferably, such a protein polypeptide has an amino acid sequence that is at least 85%, preferably 90%, and most preferably 95% or 99% identical to the amino acid sequence of the naturally-occurring (native) protein from which it is derived.

[0050] Polypeptide molecules are said to have an “amino terminus” (N-terminus) and a “carboxy terminus” (C-terminus) because peptide linkages occur between the backbone amino group of a first amino acid residue and the backbone carboxyl group of a second amino acid residue. The terms “N-terminal” and “C-terminal” in reference to polypeptide sequences refer to regions of polypeptides including portions of the N-terminal and C-terminal regions of the polypeptide, respectively. A sequence that includes a portion of the N-terminal region of polypeptide includes amino acids predominantly from the N-terminal half of the polypeptide chain, but is not limited to such sequences. For example, an N-terminal sequence may include an interior portion of the polypeptide sequence including bases from both the N-terminal and C-terminal halves of the polypeptide. The same applies to C-terminal regions. N-terminal and C-terminal regions may, but need not, include the amino acid defining the ultimate N-terminus and C-terminus of the polypeptide, respectively.

[0051] The term “wild type” as used herein, refers to a gene or gene product that has the characteristics of that gene
or gene product isolated from a naturally occurring source. A wild type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "wild type" form of the gene. In contrast, the term "mutant" refers to a gene or gene product that displays modifications in sequence and/or functional properties (i.e., altered characteristics) when compared to the wild type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild type gene or gene product.

The terms "complementary" or "complementarity" are used in reference to a sequence of nucleotides related by the base-pairing rules. For example, for the sequence 5'-A-G-T 3', is complementary to the sequence 3'-T-C-A 5'. Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon hybridization of nucleic acids.

The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule expressed from a recombinant DNA molecule. In contrast, the term "native protein" is used herein to indicate a protein isolated from a naturally occurring (i.e., nonrecombinant) source. Molecular biological techniques may be used to produce a recombinant form of a protein with identical properties as compared to the native form of the protein.

The terms "fusion protein" and "fusion partner" refer to a chimeric protein containing the protein of interest (e.g., luciferase) joined to an exogenous protein fragment (e.g., a fusion partner which consists of a non-luciferase protein). The fusion partner may enhance the solubility of the protein as expressed in a host cell, may, for example, provide an affinity tag to allow purification of the recombinant fusion protein from the host cell or culture supernatant, or both. If desired, the fusion partner may be removed from the protein of interest by a variety of enzymatic or chemical means known to the art.

The terms "cell," "cell line," "host cell," as used herein, are used interchangeably, and all such designations include progeny or potential progeny of these designations. By "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced a DNA molecule comprising a synthetic gene. Optionally, a synthetic gene of the invention may be introduced into a suitable cell line so as to create a stably-transfected cell line capable of producing the protein or polypeptide encoded by the synthetic gene. Vectors, cells, and methods for constructing such cell lines are well known in the art, e.g. in Ausubel, et al. (infra). The words "transformants" or "transformed cells" include the primary transformed cells derived from the originally transformed cell without regard to the number of transfers. All progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Nonetheless, mutant progeny that have the same functionality as screened for in the originally transformed cell are included in the definition of transformants.

Nucleic acids are known to contain different types of mutations. A "point" mutation refers to an alteration in the sequence of a nucleotide at a single base position from the wild type sequence. Mutations may also refer to insertion or deletion of one or more bases, so that the nucleic acid sequence differs from the wild-type sequence.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). Homology is often measured using sequence analysis software (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, insertions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

A "partially complementary" sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous to a target under conditions of low stringency. This is to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In this case, in the absence of non-specific binding, the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or a genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described herein.

"Probe" refers to an oligonucleotide designed to be sufficiently complementary to a sequence in a denatured nucleic acid to be probed (in relation to its length) to be bound under selected stringency conditions.

"Hybridization" and "binding" in the context of probes and denatured melted nucleic acid are used interchangeably. Probes which are hybridized or bound to denatured nucleic acid are base paired to complementary sequences in the polynucleotide. Whether or not a particular probe remains base paired with the polynucleotide depends on the degree of complementarity, the length of the probe, and the stringency of the binding conditions. The higher the stringency, the higher must be the degree of complementarity and/or the longer the probe.

The term "hybridization" is used in reference to the pairing of complementary nucleic acid strands. Hybridiza-
tion and the strength of hybridization (i.e., the strength of the association between nucleic acid strands) is impacted by many factors well known in the art including the degree of complementarity between the nucleic acids, stringency of the conditions involved affected by such conditions as the concentration of salts, the Tm (melting temperature) of the formed hybrid, the presence of other components (e.g., the presence or absence of polyethylene glycol), the molarity of the hybridizing strands and the G-C content of the nucleic acid strands.

[0063] The term “stringency” is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds, under which nucleic acid hybridizations are conducted. With “high stringency” conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of “medium” or “low” stringency are often required when it is desired that nucleic acids which are not completely complementary to one another be hybridized or annealed together. The art knows well that numerous equivalent conditions can be employed to comprise medium or low stringency conditions. The choice of hybridization conditions is generally evident to one skilled in the art and is usually guided by the purpose of the hybridization, the type of hybridization (DNA-DNA or DNA-RNA), and the level of desired relatedness between the sequences (e.g., Sarnbrook et al., 1989; Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C., 1985, for a general discussion of the methods).

[0064] The stability of nucleic acid duplexes is known to decrease with an increased number of mismatched bases, and further to be decreased to a greater or lesser degree depending on the relative positions of mismatches in the hybrid duplexes. Thus, the stringency of hybridization can be used to maximize or minimize stabilization of such duplexes. Hybridization stringency can be altered by: adjusting the temperature of hybridization; adjusting the percentage of helix destabilizing agents, such as formamide, in the hybridization mix; and adjusting the temperature and/or salt concentration of the wash solutions. For filter hybridizations, the final stringency of hybridizations often is determined by the salt concentration and/or temperature used for the post-hybridization washes.

[0065] “High stringency conditions” when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4.H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5x Denhardt’s reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1xSSPE, 1.0% SDS at 23°C when a probe of about 500 nucleotides in length is employed.

[0066] “Medium stringency conditions” when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4.H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5x Denhardt’s reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1xSSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

[0067] “Low stringency conditions” comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5xSSPE (43.8 g/l NaCl, 6.9 g/l NaH2PO4.H2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5x Denhardt’s reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5xSSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

[0068] The term “Tm”, or that is used in reference to the “melting temperature”. The melting temperature is the temperature at which 50% of a population of double-stranded nucleic acid molecules becomes dissociated into single strands. The equation for calculating the Tm of nucleic acids is well-known in the art. The Tm of a hybrid nucleic acid is often estimated using a formula adopted from hybridization assays in 1 M salt, and commonly used for calculating Tm for PCR primers: 

\[
T_m = 81.5 + 0.41 \times \text{(%GC)}
\]

(A. R. Newton et al., PCR. 2nd Ed., Springer-Verlag (New York, 1997), p. 24). This formula was found to be accurate for primers longer than 20 nucleotides. Another similar estimate of the Tm value may be calculated by the equation: 

\[
T_m = 81.5 + 0.41 \times \text{(%GC)}
\]

when a nucleic acid is in aqueous solution at 1 M NaCl. (e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization, 1985). Other more sophisticated computations exist in the art which take structural as well as sequence characteristics into account for the calculation of the Tm. A calculated Tm is merely an estimate; the optimum temperature is commonly determined empirically.

[0069] The term “isolated” when used in relation to a nucleic acid, as in “isolated oligonucleotide” or “isolated polynucleotide” refers to a nucleic acid sequence that is isolated and separated from at least one contaminant with which it is ordinarily associated in its source. Thus, an isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids (e.g., DNA and RNA) are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences (e.g., a specific mRNA sequence encoding a specific protein), are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid includes, by way of example, such nucleic acid in cells ordinarily expressing that nucleic acid where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide contains at a minimum, the sense or coding strand (i.e., the oligonucleotide may single-stranded), but may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded).

[0070] The term “isolated” when used in relation to a polypeptide, as in “isolated protein” or “isolated polypeptide” refers to a polypeptide that is isolated and separated from at least one contaminant with which it is ordinarily associated in its source. Thus, an isolated polypeptide is present in a form or setting that is different from that in
which it is found in nature. In contrast, non-isolated polypeptides (e.g., proteins and enzymes) are found in the state they exist in nature.

[0071] The term “purified” or “to purify” means the result of any process that removes some of a contaminant from the component of interest, such as a protein or nucleic acid. The percent of a purified component is thereby increased in the sample.

[0072] The term “operably linked” as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of sequences encoding amino acids in such a manner that a functional (e.g., enzymatically active, capable of binding to a binding partner, capable of inhibiting, etc.) protein or polypeptide is produced.

[0073] The term “recombinant DNA molecule” means a hybrid DNA sequence comprising at least two nucleotide sequences not normally found together in nature.

[0074] The term “vector” is used in reference to nucleic acid molecules into which fragments of DNA may be inserted or cloned and can be used to transfer DNA segment(s) into a cell and capable of replication in a cell. Vectors may be derived from plasmids, bacteriophages, viruses, cosmid, and the like.

[0075] The terms “recombinant vector” and “expression vector” as used herein refer to DNA or RNA sequences containing a desired coding sequence and appropriate DNA or RNA sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Prokaryotic expression vectors include a promoter, a ribosome binding site, an origin of replication for autonomous replication in a host cell and possibly other sequences, e.g. an optional operator sequence, optional restriction enzyme sites. A promoter is defined as a DNA sequence that directs RNA polymerase to bind to DNA and to initiate RNA synthesis. Eukaryotic expression vectors include a promoter, optionally a polyadenylation signal and optionally an enhancer sequence.

[0076] The term “a polynucleotide having a nucleotide sequence encoding a gene,” means a nucleic acid sequence comprising the coding region of a gene, or in other words the nucleic acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (i.e., the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, etc. may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, etc. In further embodiments, the coding region may contain a combination of both endogenous and exogenous control elements.

[0077] The term “transcription regulatory element” or “transcription regulatory sequence” refers to a genetic element or sequence that controls some aspect of the expression of nucleic acid sequence(s). For example, a promoter is a regulatory element that facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements include, but are not limited to, transcription factor binding sites, splicing signals, polyadenylation signals, termination signals and enhancer elements.

[0078] Transcriptional control signals in eukaryotes comprise “promoter” and “enhancer” elements. Promoters and enhancers consist of short arrays of DNA sequences that interact specifically with cellular proteins involved in transcription (Maniatis et al., 1987). Promoter and enhancer elements have been isolated from a variety of eukaryotic sources including genes in yeast, insect and mammalian cells. Promoter and enhancer elements have also been isolated from viruses and analogous control elements, such as promoters, are also found in prokaryotes. The selection of a particular promoter and enhancer depends on the cell type used to express the protein of interest. Some eukaryotic promoters and enhancers have a broad host range while others are functional in a limited subset of cell types (for review, see Voss et al., 1986; and Maniatis et al., 1987). For example, the SV40 early gene enhancer is very active in a wide variety of cell types from many mammalian species and has been widely used for the expression of proteins in mammalian cells (Dijkema et al., 1985). Two other examples of promoter/enhancer elements active in a broad range of mammalian cell types are those from the human elongation factor 1 gene (Uetsuki et al., 1989; Kim et al., 1990; and Mizushima and Nagata, 1990) and the long terminal repeats of the Rous sarcoma virus (Gorman et al., 1982); and the human cytomegalovirus (Boshart et al., 1985).

[0079] The term “promoter/enhancer” denotes a segment of DNA containing sequences capable of providing both promoter and enhancer functions (i.e., the functions provided by a promoter element and an enhancer element as described above). For example, the long terminal repeats of retroviruses contain both promoter and enhancer functions. The enhancer/promoter may be “endogenous” or “exogenous” or “heterologous.” An “endogenous” enhancer/promoter is one that is naturally linked with a given gene in the genome. An “exogenous” or “heterologous” enhancer/promoter is one that is placed in juxtaposition to a gene by means of genetic manipulation (i.e., molecular biological techniques) such that transcription of the gene is directed by the linked enhancer/promoter.

[0080] The presence of “splicing signals” on an expression vector often results in higher levels of expression of the recombinant transcript in eukaryotic host cells. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York, 1989, pp. 16.7-16.8). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

[0081] Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term
“poly(A) site” or “poly(A) sequence” as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable, as transcripts lacking a poly(A) tail are unstable and are rapidly degraded. The poly(A) signal utilized in an expression vector may be “heterologous” or “endogenous.” An endogenous poly(A) signal is one that is found naturally at the 3’ end of the coding region of a given gene in the genome. A heterologous poly(A) signal is one which has been isolated from one gene and positioned 3’ to another gene. A commonly used heterologous poly(A) signal is the SV40 poly(A) signal. The SV40 poly(A) signal is contained on a 237 bp BamH I/Bgl I restriction fragment and directs both termination and polyadenylation (Sambrook, supra, at 16.6-16.7).

[0082] Eukaryotic expression vectors may also contain “viral replicons” or “viral origins of replication.” Viral replicons are viral DNA sequences which allow for the extrachromosomal replication of a vector in a host cell expressing the appropriate replication factors. Vectors containing either the SV40 or polyoma virus origin of replication replicate to high copy number (up to 10^6 copies/cell) in cells that express the appropriate viral T antigen. In contrast, vectors containing the replicons from bovine papillomavirus or Epstein-Barr virus replicate extrachromosomally at low copy number (about 100 copies/cell).

[0083] The term “in vitro” refers to an artificial environment and to processes or reactions that occur within an artificial environment. In vitro environments include, but are not limited to, test tubes and cell lysates. The term “in situ” refers to cell culture. The term “in vivo” refers to the natural environment (e.g., an animal or a cell) and to processes or reaction that occur within a natural environment.

[0084] The term “expression system” refers to any assay or system for determining (e.g., detecting) the expression of a gene of interest. Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems may be used. A wide range of suitable mammalian cells are available from a wide range of source (e.g., the American Type Culture Collection, Rockland, Md.). The method of transformation or transfection and the choice of expression vehicle will depend on the host system selected. Transformation and transfection methods are described, e.g., in Ausubel, et al., Current Protocols in Molecular Biology. John Wiley & Sons, New York. 1992. Expression systems include in vitro gene expression assays where a gene of interest (e.g., a reporter gene) is linked to a regulatory sequence and the expression of the gene is monitored following treatment with an agent that inhibits or induces expression of the gene. Detection of gene expression can be through any suitable means including, but not limited to, detection of expressed mRNA or protein (e.g., a detectable product of a reporter gene) or through a detectable change in the phenotype of a cell expressing the gene of interest. Expression systems may also comprise assays where a cleavage event or other nucleic acid or cellular change is detected.

[0085] The term “enzyme” refers to molecules or molecular aggregates that are responsible for catalyzing chemical and biological reactions. Such molecules are typically proteins, but can also comprise short peptides, RNAs, ribozymes, antibodies, and other molecules. A molecule that catalyzes chemical and biological reactions is referred to as “having enzyme activity” or “having catalytic activity.”

[0086] All amino acid residues identified herein are in the natural L-configuration. In keeping with standard polypeptide nomenclature (see J. Biol. Chem., 243, 3557 (1969)), abbreviations for amino acid residues are as shown in the following Table of Correspondence.

<table>
<thead>
<tr>
<th>TABLE OF CORRESPONDENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Letter</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>F</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>I</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>K</td>
</tr>
<tr>
<td>H</td>
</tr>
<tr>
<td>Q</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>W</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>

[0087] The term “sequence homology” means the proportion of base matches between two nucleic acid sequences or the proportion of amino acid matches between two amino acid sequences. When sequence homology is expressed as a percentage, e.g., 50%, the percentage denotes the proportion of matches over the length of sequence from one sequence that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of 15 bases or less are usually used, 6 bases or less are preferred with 2 bases or less more preferred. When using oligonucleotides as probes or treatments, the sequence homology between the target nucleic acid and the oligonucleotide sequence is generally not less than 17 target base matches out of 20 possible oligonucleotide base pair matches (85%); preferably not less than 9 matches out of 10 possible base pair matches (90%), and more preferably not less than 19 matches out of 20 possible base pair matches (95%).

[0088] Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred. Alternatively and preferably, two protein sequences (or polypeptide sequences derived from them) of at least 100 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of at least 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M. O., in Atlas of Protein Sequence and Structure,
The following terms are used to describe the sequence relationships between two or more nucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar to the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988); the local homology algorithm of Smith and Waterman (1981); the homology alignment algorithm of Needleman and Wunsch (1970); the search for similarity method of Pearson and Lipman (1988); and the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins et al. (1988); Higgins et al. (1989); Corpet et al. (1988); Huang et al. (1992); and Pearson et al. (1994). The ALIGN program is based on the algorithm of Myers and Miller, supra. The BLAST programs of Altschul et al. (1990), are based on the algorithm of Karlin and Altschul supra. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul et al. (1997). Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul et al., supra. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g. BLASTN for nucleotide sequences, BLASTX for proteins) can be used. See http://www.ncbi.nlm.nih.gov. Alignment may also be performed manually by inspection.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) for the stated proportion of nucleotides over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 90%, preferably at least 95%, more preferably at least 98%, up to about 99%, even more preferably at least 99%, more usually at least 95%, sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 20-50 nucleotides, and preferably at least 300 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a larger sequence.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least about 85% sequence identity, preferably at least about 90% sequence identity, more preferably at least about 95% sequence identity, and most preferably at least about 98% sequence identity.

The Synthetic Nucleic Acid Molecules and Methods of the Invention

The invention provides compositions comprising synthetic nucleic acid molecules, as well as methods for preparing those molecules which yield synthetic nucleic acid molecules that are efficiently expressed as a polypeptide or protein with desirable characteristics including reduced inappropriate or unintended transcription characteristics when expressed in a particular cell type.

Natural selection is the hypothesis that genotype-environment interactions occurring at the phenotypic level lead to differential reproductive success of individuals and hence to modification of the gene pool of a population. It is generally accepted that the amino acid sequence of a protein found in nature has undergone optimization by natural selection. However, amino acids exist within the sequence of
a protein that do not contribute significantly to the activity of the protein and these amino acids can be changed to other amino acids with little or no consequence. Furthermore, a protein may be useful outside its natural environment for purposes that differ from the conditions of its natural selection. In these circumstances, the amino acid sequence can be synthetically altered to better adapt the protein for its utility in various applications.

Likewise, the nucleic acid sequence that encodes a protein is also optimized by natural selection. The relationship between coding DNA and its transcribed RNA is such that any change to the DNA affects the resulting RNA. Thus, natural selection works on both molecules simultaneously. However, this relationship does not exist between nucleic acids and proteins. Because multiple codons encode the same amino acid, many different nucleotide sequences can encode an identical protein. A specific protein composed of 500 amino acids can theoretically be encoded by more than $10^{150}$ different nucleic acid sequences.

Natural selection acts on nucleic acids to achieve proper encoding of the corresponding protein. Presumably, other properties of nucleic acid molecules are also acted upon by natural selection. These properties include codon usage frequency, RNA secondary structure, the efficiency of intron splicing, and interactions with transcription factors or other nucleic acid binding proteins. These other properties may alter the efficiency of protein translation and the resulting phenotype. Because of the redundant nature of the genetic code, these other attributes can be optimized by natural selection without altering the corresponding amino acid sequence.

Under some conditions, it is useful to synthetically alter the natural nucleotide sequence encoding a protein to better adapt the protein for alternative applications. A common example is to alter the codon usage frequency of a gene when it is expressed in a foreign host. Although redundancy in the genetic code allows amino acids to be encoded by multiple codons, different organisms favor some codons over others. The codon usage frequencies tend to differ most for organisms with widely separated evolutionary histories. It has been found that when transferring genes between evolutionarily distant organisms, the efficiency of protein translation can be substantially increased by adjusting the codon usage frequency (see U.S. Pat. Nos. 5,096,825, 5,670,356 and 5,874,304).

Because of the need for evolutionary distance, the codon usage of reporter genes often does not correspond to the optimal codon usage of the experimental cells. Examples include $\beta$-galactosidase ($\beta$-gal) and chloramphenicol acetyltransferase (cat) reporter genes that are derived from $E. coli$ and are commonly used in mammalian cells; the $\beta$-glucuronidase (gus) reporter gene that is derived from $E. coli$ and commonly used in plant cells; the firefly luciferase (luc) reporter gene that is derived from an insect and commonly used in plant and mammalian cells; and the Renilla luciferase, and green fluorescent protein (gfp) reporter genes which are derived from coelenterates and are commonly used in plant and mammalian cells. To achieve sensitive quantification of reporter gene expression, the activity of the gene product must not be endogenous to the experimental host cells. Thus, reporter genes are usually selected from organisms having unique and distinctive phenotypes. Consequently, these organisms often have widely separated evolutionary histories from the experimental host cells.

Previously, to create genes having a more optimal codon usage frequency but still encoding the same gene product, a synthetic nucleic acid sequence was made by replacing existing codons with codons that were generally more favorable to the experimental host cell (see U.S. Pat. Nos. 5,096,825, 5,670,356 and 5,874,304.) The result was a net improvement in codon usage frequency of the synthetic gene. However, the optimization of other attributes was not considered and so these synthetic genes likely did not reflect genes optimized by natural selection.

In particular, improvements in codon usage frequency are intended only for optimization of a RNA sequence based on its role in translation into a protein. Thus, previously described methods did not address how the sequence of a synthetic gene affects the role of DNA in transcription into RNA. Most notably, consideration had not been given as to how transcription factors may interact with the synthetic DNA and consequently modulate or otherwise influence gene transcription. For genes found in nature, the DNA would be optimally transcribed by the native host cell and would yield an RNA that encodes a properly folded gene product. In contrast, synthetic genes have previously not been optimized for transcriptional characteristics. Rather, this property has been ignored or left to chance.

This concern is important for all genes, but particularly important for reporter genes, which are most commonly used to quantitate transcriptional behavior in the experimental host cells. Hundreds of transcription factors have been identified in different cell types under different physiological conditions, and likely more exist but have not yet been identified. All of these transcription factors can influence the transcription of an introduced gene. A useful synthetic reporter gene of the invention has a minimal risk of influencing or perturbing intrinsic transcriptional characteristics of the host cell because the structure of that gene has been altered. A particularly useful synthetic reporter gene will have desirable characteristics under a new set and/or a wide variety of experimental conditions. To best achieve these characteristics, the structure of the synthetic gene should have minimal potential for interacting with transcription factors within a broad range of host cells and physiological conditions. Minimizing potential interactions between a reporter gene and a host cell’s endogenous transcription factors increases the value of a reporter gene by reducing the risk of inappropriate transcriptional characteristics of the gene within a particular experiment, increasing applicability of the gene in various environments, and increasing the acceptance of the resulting experimental data.

In contrast, a reporter gene comprising a native nucleotide sequence, based on a genomic or cDNA clone from the original host organism, may interact with transcription factors when expressed in an exogenous host. This risk stems from two circumstances. First, the native nucleotide sequence contains sequences that were optimized through natural selection to influence gene transcription within the native host organism. However, these sequences might also influence transcription when the gene is expressed in exogenous hosts, i.e., out of context, thus interfering with its performance as a reporter gene. Second, the nucleotide sequence may inadvertently interact with transcription fac-
tors that were not present in the native host organism, and thus did not participate in its natural selection. The probability of such inadvertent interactions increases with greater evolutionary separation between the experimental cells and the native organism of the reporter gene.

[0105] These potential interactions with transcription factors would likely be disrupted when using a synthetic reporter gene having alterations in codon usage frequency. However, a synthetic reporter gene sequence, designed by choosing codons based only on codon usage frequency, is likely to contain other unintended transcription factor binding sites since the synthetic gene has not been subjected to the benefit of natural selection to correct inappropriate transcriptional activities. Inadvertent interactions with transcription factors could also occur whenever the encoded amino acid sequence is artificially altered, e.g., to introduce amino acid substitutions. Similarly, these changes have not been subjected to natural selection, and thus may exhibit undesired characteristics.

[0106] Thus, the invention provides a method for preparing synthetic nucleic acid sequences that reduce the risk of undesirable interactions of the nucleic acid with transcription factors when expressed in a particular host cell, thereby reducing inappropriate or unintended transcriptional characteristics. Preferably, the method yields synthetic genes containing improved codon usage frequencies for a particular host cell and with a reduced occurrence of transcription factor binding sites. The invention also provides a method of preparing synthetic genes containing improved codon usage frequencies with a reduced occurrence of transcription factor binding sites and additional beneficial structural attributes. Such additional attributes include the absence of inappropriate RNA splicing junctions, poly(A) addition signals, undesirable restriction sites, ribosomal binding sites, and secondary structural motifs such as hairpin loops.

[0107] Also provided is a method for preparing two synthetic genes encoding the same or highly similar proteins ("codon distinct" versions). Preferably, the two synthetic genes have a reduced ability to hybridize to a common polynucleotide probe sequence, or have a reduced risk of recombining when present together in living cells. To detect recombination, PCR amplification of the reporter sequences using primers complementary to flanking sequences and sequencing of the amplified sequences may be employed.

[0108] To select codons for the synthetic nucleic acid molecules of the invention, preferred codons have a relatively high codon usage frequency in a selected host cell, and their introduction results in the introduction of relatively few transcription factor binding sites, relatively few other undesirable structural attributes, and optionally a characteristic that distinguishes the synthetic gene from another gene encoding a highly similar protein. Thus, the synthetic nucleic acid product obtained by the method of the invention is a synthetic gene with improved level of expression due to improved codon usage frequency, a reduced risk of inappropriate transcriptional behavior due to a reduced number of undesirable transcription regulatory sequences, and optionally any additional characteristic due to other criteria that may be employed to select the synthetic sequence.

[0109] The invention may be employed with any nucleic acid sequence, e.g., a native sequence such as a cDNA or one which has been manipulated in vitro, e.g., to introduce specific alterations such as the introduction or removal of a restriction enzyme recognition site, the alteration of a codon to encode a different amino acid or to encode a fusion protein, or to alter GC or AT content (% of composition) of nucleic acid molecules. Moreover, the method of the invention is useful with any gene, but particularly useful for reporter genes as well as other genes associated with the expression of reporter genes, such as selectable markers. Preferred genes include, but are not limited to, those encoding lactamase (β-gal), neomycin resistance (Neo), CAT, GUS, galactopyranoside, GFP, xylosidase, thymidine kinase, arabinosidase and the like. As used herein, a "marker gene" or "reporter gene" is a gene that imparts a distinct phenotype to cells expressing the gene and thus permits cells having the gene to be distinguished from cells that do not have the gene. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can 'select' for by chemical means, i.e., through the use of a selective agent (e.g., a herbicide, antibiotic, or the like), or whether it is simply a "reporter" trait that one can identify through observation or testing, i.e., by 'screening'. Elements of the present disclosure are exemplified in detail through the use of particular marker genes. Of course, many examples of suitable marker genes or reporter genes are known to the art and can be employed in the practice of the invention. Therefore, it will be understood that the following discussion is exemplary rather than exhaustive. In light of the techniques disclosed herein and the general recombinant techniques which are known in the art, the present invention renders possible the alteration of any gene.

[0110] Exemplary marker genes include, but are not limited to, a neo gene, a β-gal gene, a gus gene, a cat gene, a gpt gene, a hyg gene, a hisD1 gene, a ble gene, a mph gene, a bar gene, a nitrilase gene, a mutant acetalactate synthase gene (ALS) or acetolactate synthase gene (AAS), a methotrexate-resistant dhfr gene, a dalapon dehalogenase gene, a mutated anthranilate synthase gene that confers resistance to 5-methyl tryptophan (WO 97/26366), an R locus gene, a β-lactamase gene, a xylE gene, an α-amylase gene, a tyrosinase gene, a luciferase (luc) gene, (e.g., a Renilla reniformis luciferase gene, a firefly luciferase gene, or a click beetle luciferase (Pyrophorus plagiophthalamus) gene), an aequorin gene, or a green fluorescent protein gene. Included within the terms selectable or screenable marker genes are also genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, e.g., by ELISA, and proteins that are inserted or trapped in the cell membrane.

[0111] The method of the invention can be performed by, although it is not limited to, a recursive process. The process includes assigning preferred codons to each amino acid in a target molecule, e.g., a native nucleotide sequence, based on codon usage in a particular species, identifying potential transcription regulatory sequences such as transcription factor binding sites in the nucleic acid sequence having preferred codons, e.g., using a database of such binding sites, optionally identifying other undesirable sequences, and substituting an alternative codon (i.e., encoding the same amino
acid) at positions where undesirable transcription factor binding sites or other sequences occur. For codon distinct versions, alternative preferred codons are substituted in each version. If necessary, the identification and elimination of potential transcription factor or other undesirable sequences can be repeated until a nucleotide sequence is achieved containing a maximum number of preferred codons and a minimum number of undesired sequences including transcription regulatory sequences or other undesirable sequences. Also, optionally, desired sequences, e.g., restriction enzyme recognition sites, can be introduced. After a synthetic nucleic acid molecule is designed and constructed, its properties relative to the parent nucleic acid sequence can be determined by methods well known to the art. For example, the expression of the synthetic and target nucleic acid molecules in a series of vectors in a particular cell can be compared.

Thus, generally, the method of the invention comprises identifying a target nucleic acid sequence, such as a vector backbone, a reporter gene or a selectable marker gene, and a host cell of interest, for example, a plant (dicot or monocot), fungus, yeast or mammalian cell. Preferred host cells are mammalian host cells such as CHO, COS, 293, Hela, CV-1 and NIH313 cells. Based on preferred codon usage in the host cell(s) and, optionally, low codon usage in the host cell(s), e.g., high usage mammalian codons and low usage E. coli and mammalian codons, codons to be replaced are determined. For codon distinct versions of two synthetic nucleic acid molecules, alternative preferred codons are introduced to each version. Thus, for amino acids having more than two codons, one preferred codon is introduced to one version and another preferred codon is introduced to the other version. For amino acids having six codons, the two codons with the largest number of mismatched bases are identified and one is introduced to one version and the other codon is introduced to the other version. Concurrent, subsequent or prior to selecting codons to be replaced, desired and undesired sequences, such as undesired transcriptional regulatory sequences, in the target sequence are identified. These sequences can be identified using databases and software such as EPM, NNPD, REBASE, TRANSFAC, TESS, GenePro, MAR (www.ncbi.org/MAR-search) and BCM Gene Finder, further described herein. After the sequences are identified, the modification(s) are introduced. Once a desired synthetic nucleic acid sequence is obtained, it can be prepared by methods well known to the art (such as PCR with overlapping primers), and its structural and functional properties compared to the target nucleic acid sequence, including, but not limited to, percent homology, presence or absence of certain sequences, for example, restriction sites, percent of codons changed (such as an increased or decreased usage of certain codons) and expression rates.

As described below, the method was used to create synthetic reporter genes encoding *Renilla reniformis* luciferase, and two click beetle luciferases (one emitting green light and the other emitting red light). For both systems, the synthetic genes support much greater levels of expression than the corresponding native or parent genes for the protein. In addition, the native and parent genes demonstrated anomalous transcription characteristics when expressed in mammalian cells, which were not evident in the synthetic genes. In particular, basal expression of the native or parent genes is relatively high. Furthermore, the expression is induced to very high levels by an enhancer sequence in the absence of known promoters. The synthetic genes show lower basal expression and do not show the anomalous enhancer behavior. Presumably, the enhancer is activating transcriptional elements found in the native genes that are absent in the synthetic genes. The results clearly show that the synthetic nucleic acid sequences exhibit superior performance as reporter genes.

Exemplary Uses of the Molecules of the Invention

The synthetic genes of the invention preferably encode the same proteins as their native counterparts (or nearly so), but have improved codon usage while being largely devoid of known transcription regulatory elements in the coding region. (It is recognized that a small number of amino acid changes may be desired to enhance a property of the native counterpart protein, e.g., to enhance luminescence of a luciferase.) This increases the level of expression of the protein the synthetic gene encodes and reduces the risk of anomalous expression of the protein. For example, studies of many important events of gene regulation, which may be mediated by weak promoters, are limited by insufficient reporter signals from inadequate expression of the reporter proteins. The synthetic luciferase genes described herein permit detection of weak promoter activity because of the large increase in level of expression, which enables increased detection sensitivity. Also, the use of some selectable markers may be limited by the expression of that marker in an exogenous cell. Thus, synthetic selectable marker genes which have improved codon usage for that cell, and have a decrease in other undesirable sequences, (e.g., transcription factor binding sites), can permit the use of those markers in cells that otherwise were undesirable as hosts for those markers.

Promoter crosstalk is another concern when a co-reporter gene is used to normalize transfection efficiencies. With the enhanced expression of synthetic genes, the amount of DNA containing strong promoters can be reduced, or DNA containing weaker promoters can be employed, to drive the expression of the co-reporter. In addition, there may be a reduction in the background expression from the synthetic reporter genes of the invention. This characteristic makes synthetic reporter genes more desirable by minimizing the sporadic expression from the genes and reducing the interference resulting from other regulatory pathways.

The use of reporter genes in imaging systems, which can be used for in vivo biological studies or drug screening, is another use for the synthetic genes of the invention. Due to their increased level of expression, the protein encoded by a synthetic gene is more readily detectable by an imaging system. In fact, using a synthetic *Renilla* luciferase gene, luminescence in transfected CHO cells was detected visually without the aid of instrumentation.

In addition, the synthetic genes may be used to express fusion proteins, for example fusions with secretion leader sequences or cellular localization sequences, to study transcription in difficult-to-transfect cells such as primary cells, and/or to improve the analysis of regulatory pathways and genetic elements. Other uses include, but are not limited to, the detection of rare events that require extreme sensitivity (e.g., studying RNA recoding), use with IRIS, to improve the efficiency of in vitro translation or in vitro
transcription-translation coupled systems such as TNT (Promega Corp., Madison, Wis.), study of reporters optimized to different host organisms (e.g., plants, fungus, and the like), use of multiple genes as co-reporters to monitor drug toxicity, as reporter molecules in multiwell assays, and as reporter molecules in drug screening with the advantage of minimizing possible interference of reporter signal by different signal transduction pathways and other regulatory mechanisms.

Additionally, uses for the nucleic acid molecules of the invention include fluorescence activated cell sorting (FACS), fluorescent microscopy, to detect and/or measure the level of gene expression in vitro and in vivo, (e.g., to determine promoter strength), subcellular localization or targeting (fusion protein), as a marker, in calibration, in a kit, (e.g., for dual assays), for in vivo imaging, to analyze regulatory pathways and genetic elements, and in multi-well formats.

With respect to synthetic DNA encoding luciferases, the use of synthetic click beetle luciferases provides advantages such as the measurement of dual reporters. As *Renilla* luciferase is better suited for in vivo imaging (because it does not depend on ATP or Mg²⁺ for reaction, unlike firefly luciferase, and because coelenterazine is more permeable to the cell membrane than luciferin), the synthetic *Renilla* luciferase gene can be employed in vivo. Further, the synthetic *Renilla* luciferase has improved fidelity and sensitivity in dual luciferase assays, e.g., for biological analysis or in drug screening platform.

Demonstration of the Invention Using Luciferase Genes

The reporter genes for click beetle luciferase and *Renilla* luciferase were used to demonstrate the invention because the reaction catalyzed by the protein they encode are significantly easier to quantify than the product of most genes. However, for the purposes of demonstrating the present invention they represent genes in general.

Although the click beetle luciferase and *Renilla* luciferase genes share the name “luciferase”, this should not be interpreted to mean that they originate from the same family of genes. The two luciferase proteins are evolutionarily distinct; they have fundamentally different traits and physical structures, they use vastly different substrates (FIG. 17), and they evolved from completely different families of genes. The click beetle luciferase is 61 kD in size, uses luciferin as a substrate and evolved from the CoA synthetases. The *Renilla* luciferase originates from the sea pansy *Renilla reniformis*, is 35 kD in size, uses coelenterazine as a substrate and evolved from the COβ hydrolases. The only shared trait of these two enzymes is that the reaction they catalyze results in light output. They are no more similar for resulting in light output than any other two enzymes would be, for example, simply because the reaction they catalyze results in heat.

Bioluminescence is the light produced in certain organisms as a result of luciferase-mediated oxidation reactions. The luciferase genes, e.g., the genes from luminous beetles, sea pansy, and, in particular, the luciferase from *Photinus pyralis* (the common firefly of North America), are currently the most popular luminescent reporter genes. Reference is made to Bronstein et al. (1994) for a review of luminescent reporter gene assays and to Wood (1995) for a review of the evolution of beetle bioluminescence. See FIG. 17 for an illustration of the reactions catalyzed by each of firefly and click beetle luciferases (17A) and *Renilla* luciferase (17B).

Firefly luciferase and *Renilla* luciferase are highly valuable as genetic reporters due to the convenience, sensitivity and linear range of the luminescence assay. Today, luciferase is used in virtually every type of experimental biological system, including, but not limited to, prokaryotic and eukaryotic cell culture, transgenic plants and animals, and cell-free expression systems. The firefly luciferase enzyme is derived from a specific North American beetle, *Photinus pyralis*. The firefly luciferase enzyme and the click beetle luciferase enzyme are monomeric proteins (61 kDa) which generate light through monoxygenation of beetle luciferin utilizing ATP and O₂ (FIG. 17A). The *Renilla* luciferase is derived from the sea pansy *Renilla reniformis*. The *Renilla* luciferase enzyme is a 36 kDa monomeric protein that utilizes O₂ and coelenterazine to generate light (FIG. 17B).

The gene encoding firefly luciferase was cloned from *Photinus pyralis*, and demonstrated to produce active enzyme in *E. coli* (de Wet et al., 1987). The cDNA encoding firefly luciferase (luc) continues to gain favor as the gene of choice for reporting genetic activity in animal, plant and microbial cells. The firefly luciferase reaction, modified by the addition of CoA to produce persistent light emission, provides an extremely sensitive and rapid in vitro assay for quantifying firefly luciferase expression in small samples of transfected cells or tissues.

To use firefly luciferase or click beetle luciferase as a genetic reporter, extracts of cells expressing the luciferase are mixed with substrates (beetle luciferin, Mg²⁺ ATP, and O₂), and luminescence is measured immediately. The assay is very rapid and sensitive, providing gene expression data with little effort. The conventional firefly luciferase assay has been further improved by including coenzyme A in the assay reagent to yield greater enzyme turnover and thus greater luminescence intensity (Promega Luciferase Assay Reagent, Cat.# E1500, Promega Corporation, Madison, Wis.). Using this reagent, luciferase activity can be readily measured in luminometers or scintillation counters. Firefly and click beetle luciferase activity can also be detected in living cells in culture by adding luciferin to the growth medium. This in situ luminescence relies on the ability of beetle luciferin to diffuse through cellular and peroxisomal membranes and on the intracellular availability of ATP and O₂ in the cytosol and peroxisome.

Further, although reporter genes are widely used to measure transcription events, their utility can be limited by the fidelity and efficiency of reporter expression. For example, in U.S. Pat. No. 5,670,356, a firefly luciferase gene (referred to as luc+) was modified to improve the level of luciferase expression. While a higher level of expression was observed, it was not determined that higher expression had improved regulatory control.

The invention will be further described by the following nonlimiting examples.
EXAMPLE 1

Synthetic Click Beetle (RD and GR) Luciferase
Nucleic Acid Molecules

[0128] LucPpLYG is a wild-type click beetle luciferase that emits yellow-green luminescence (Wood, 1989). A mutant of LucPpLYG named YG#81-6G01 was envisioned. YG#81-6G01 lacks a phenixosome targeting signal, has a lower Km for luciferin and ATP, has increased signal stability and increased temperature stability when compared to the wild type (PCT/WO9914336). YG#81-6G01 was mutated to emit green luminescence by changing Ala at position 224 to Val (A224V is a green-shifting mutation), or to emit red luminescence by simultaneously introducing the amino acid substitutions A224H, S247H, N346I, and H348Q (red-shifting mutation set) (PCT/WO9918853)

[0129] Using YG#81-6G01 as a parent gene, two synthetic gene sequences were designed. One codes for a luciferase emitting green luminescence (GR) and one for a luciferase emitting red luminescence (RD). Both genes were designed to 1) have optimized codon usage for expression in mammalian cells, 2) have a reduced number of transcriptional regulatory sites including mammalian transcription factor binding sites, splice sites, poly(A) addition sites and promoters, as well as prokaryotic (E. coli) regulatory sites, 3) be devoid of unwanted restriction sites, e.g., those which are likely to interfere with standard cloning procedures, and 4) have a low DNA sequence identity compared to each other in order to minimize genetic rearrangements when both are present inside the same cell. In addition, desired sequences, e.g., a Kozak sequence or restriction enzyme recognition sites, may be identified and introduced.

[0130] Not all design criteria could be met equally well at the same time. The following priority was established for reduction of transcriptional regulatory sites: elimination of transcription factor (TF) binding sites received the highest priority, followed by elimination of splice sites and poly(A) addition sites, and finally prokaryotic regulatory sites. When removing regulatory sites, the strategy was to work from the lesser important to the most important to ensure that the most important changes were made first. Then the sequence was rechecked for the appearance of new lower priority sites and additional changes made as needed. Thus, the process for designing the synthetic GR and RD gene sequences, using computer programs described herein, involved 5 optionally iterative steps that are detailed below

[0131] 1. Optimized codon usage and changed A224V to create GRver1, separately changed A224H, S247H, H348Q and N346I to create RDver1. These particular amino acid changes were maintained throughout all subsequent manipulations to the sequence.

[0132] 2. Removed undesired restriction sites, prokaryotic regulatory sites, splice sites, poly(A) sites thereby creating GRver2 and RDver2.

[0133] 3. Removed transcription factor binding sites (first pass) and removed any newly created undesired sites as listed in step 2 above thereby creating GRver3 and RDver3.

[0134] 4. Removed transcription factor binding sites created by step 3 above (second pass) and removed any newly created undesired sites as listed in step 2 above thereby creating GRver4 and RDver4.

[0135] 5. Removed transcription factor binding sites created by step 4 above (third pass) and confirmed absence of sites listed in step 2 above thereby creating GRver5 and RDver5.

[0136] 6. Constructed the actual genes by PCR using synthetic oligonucleotides corresponding to fragments of GRver5 and RDver5 designed sequences (FIGS. 6 and 10) thereby creating GR7 and RD7. GR6, upon sequencing was found to have the serine residue at amino acid position 49 mutated to an asparagine and the proline at amino acid position 230 mutated to a serine (S49N, P230S). RD7, upon sequencing was found to have the histidine at amino acid position 36 mutated to a tyrosine (H36Y). These changes occurred during the PCR process.

[0137] 7. The mutations described in step 6 above (S49N, P230S for GR6 and H36Y for RD7) were reversed to create GRver5.1 and RDver5.1.

[0138] 8. RDver5.1 was further modified by changing the arginine codon at position 351 to a glycine codon (R351G) thereby creating RDver5.2 with improved spectral properties compared to RDver5.1.

[0139] 9. RDver5.2 was further mutated to increase luminescence intensity thereby creating RD156-H9 which encodes four additional amino acid changes (M21, S349T, K488T, E538V) and three silent single base changes (SEQ ID NO:18).

1. Optimize Codon Usage and Introduce Mutations Determining Luminescence Color

[0140] The starting gene sequence for this design step was YG#81-6G01 (SEQ ID NO:2).

a) Optimize Codon Usage:

[0141] The strategy was to adapt the codon usage for optimal expression in human cells and at the same time to avoid E. coli low-usagge codons. Based on these requirements, the best two codons for expression in human cells for all amino acids with more than two codons were selected (see Wada et al., 1990). In the selection of codon pairs for amino acids with six codons, the selection was biased towards pairs that have the largest number of mismatched bases to allow design of GR and RD genes with minimum sequence identity (codon distinction):

Arg: GCC/GCT Leu: CCG/CTG Ser: TCT/ACT
Thr: ACC/ACT Pro: CCA/CCT Ala: GCC/GCT
Gly: GCC/GGT Val: GTC/TTT Ile: ATC/ATT

Based on this selection of codons, two gene sequences encoding the YG#81-6G01 luciferase protein sequence were computer generated. The two genes were designed to have minimum DNA sequence identity and at the same time closely similar codon usage. To achieve this, each codon in the two genes was replaced by a codon from the limited list described above in an alternating fashion (e.g., Arg(1) is GCC in gene 1 and CGT in gene 2, Arg(2) is CGT in gene 1 and CGC in gene 2).
For subsequent steps in the design process it was anticipated that changes had to be made to this limited optimal codon selection in order to meet other design criteria, however, the following low-usage codons in mammalian cells were not used unless needed to meet criteria of higher priority:


Also, the following low-usage codons in *E. coli* were avoided when reasonable (note that 3 of these match the low-usage list for mammalian cells):

Arg: CGA/CGG/AGA/AGG  Leu: CTA  Pro: CCG  Ile: ATA

b) Introduce Mutations Determining Luminescence Color:

Into one of the two codon-optimized gene sequences was introduced the single green-shifting mutation and into the other were introduced the 4 red-shifting mutations as described above.

The two output sequences from this first design step were named GRver1 (version 1 GR) and RDver1 (version 1 RD). Their DNA sequences are 63% identical (594 mismatches), while the proteins they encode differ only by the 4 amino acids that determine luminescence color (see FIGS. 2 and 3 for an alignment of the DNA and protein sequences).

Tables 1 and 2, as an example, the codon usage for valine and leucine in human genes, the parent gene YG#81-6G01, the codon-optimized synthetic genes GRver1 and RDver1, as well as the final versions of the synthetic genes after completion of step 5 in the design process (GRver5 and RDver5). For a complete summary of the codon changes, see FIGS. 4 and 5.

### TABLE 1

<table>
<thead>
<tr>
<th>Codon</th>
<th>Human</th>
<th>Parent</th>
<th>GR ver1</th>
<th>RD ver1</th>
<th>GR ver5</th>
<th>RD ver5</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTA</td>
<td>4</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GTC</td>
<td>13</td>
<td>4</td>
<td>25</td>
<td>24</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>GTG</td>
<td>24</td>
<td>12</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>GTT</td>
<td>9</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

### TABLE 2-continued

#### Leucine

<table>
<thead>
<tr>
<th>Codon</th>
<th>Human</th>
<th>Parent</th>
<th>GR ver1</th>
<th>RD ver1</th>
<th>GR ver5</th>
<th>RD ver5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTA</td>
<td>3</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TTG</td>
<td>6</td>
<td>13</td>
<td>27</td>
<td>27</td>
<td>23</td>
<td>25</td>
</tr>
</tbody>
</table>

2. Remove Undesired Restriction Sites, Prokaryotic Regulatory Sites, Splice Sites and Poly(A) Addition Sites

a) Remove Undesired Restriction Sites:

To check for the presence and location of undesired restriction sites, the sequences of both synthetic genes were compared against a database of restriction enzyme recognition sequences (REBASE ver.7.12, http://www.neb.com/rebase) using standard sequence analysis software (GenePro ver 6.10, Riverside Scientific Ent.).

Specifically, the following restriction enzymes were classified as undesired:

- BamH I, Xho I, Sfi I, Kpn I, Sac I, Mlu I, Nhe I, Sma I, Xho I, Bgl II, Hind III, Nco I, Nar I, Xba I, Hpa I, Sal I,
- other cloning sites commonly used: EcoR I, EcoR V, Cla I,
- eight-base cutters (commonly used for complex constructs),
- BstE II (to allow N-terminal fusions),
- Xcm I (can generate A/T overhang used for T-vector cloning).

To eliminate undesired restriction sites when found in a synthetic gene, one or more codons of the synthetic gene sequence were altered in accordance with the codon optimization guidelines described in 1a above.

b) Remove Prokaryotic (*E. coli*) Regulatory Sequences:

To check for the presence and location of prokaryotic regulatory sequences, the sequences of both synthetic genes were searched for the presence of the following consensus sequences using standard sequence analysis software (GenePro):

- TATAAT (~-10 Pribnow box of promoter)
- AGGA or GGAG (ribosome binding site; only considered if paired with a methionine codon 12 or fewer bases downstream).

To eliminate such regulatory sequences when found in a synthetic gene, one or more codons of the synthetic gene at sequence were altered in accordance with the codon optimization guidelines described in 1a above.

c) Remove Splice Sites:

To check for the presence and location of splice sites, the DNA strand corresponding to the primary RNA transcript of each synthetic gene was searched for the
presence of the following consensus sequences (see Watson et al., 1983) using standard sequence analysis software (GenePro):

splice donor site: AG/GTRAGT (exon/intron), the search was performed for AGGTRAG and the lower stringency GGTRAGT;

splice acceptor site: (Y)NCAG/G (intron/exon), the search was performed with n=1.

To eliminate splice sites found in a synthetic gene, one or more codons of the synthetic gene sequence were altered in accordance with the codon optimization guidelines described in 1a above. Splice acceptor sites were generally difficult to eliminate in one gene without introducing them into the other gene because they tended to contain one of the two only Gln codons (CAG); they were removed by placing the Gln codon CAA in both genes at the expense of a slightly increased sequence identity between the two genes.

d) Remove Poly(A) Addition Sites:

To eliminate each poly(A) addition site found in a synthetic gene, one or more codons of the synthetic gene sequence were altered in accordance with the codon optimization guidelines described in 1a above. The two output sequences from this second design step were named GRver2 and RDver2. Their DNA sequences are 63% identical (590 mismatches) (FIGS. 2 and 3).

3. Remove Transcription Factor (TF) Binding Sites, then Repeat Steps 2 a-d

The starting gene sequences for this design step were GRver2 and RDver2. To check for the presence, location and identity of potential TF binding sites, the sequences of both synthetic genes were used as query sequences to search a database of transcription factor binding sites (TRANSFAC v3.2). The TRANSFAC database (http://transfac.gbf.de/TRANSFAC/index.html) holds information on gene regulatory DNA sequences (TF binding sites) and proteins (TFs) that bind to and act through them. The SITE table of TRANSFAC Release 3.2 contains 4,401 entries of individual (putative) TF binding sites (including TF binding sites in eukaryotic genes), in artificial sequences resulting from mutagenesis studies and in vitro selection procedures based on random oligonucleotide mixtures or specific theoretical considerations, and consensus binding sequences (from Faisst and Meyer, 1992).

The software tool used to locate and display these TF binding sites in the synthetic gene sequences was TESS (Transcription Element Search Software, http://agave.hggen.upenn.edu/tess/index.html). The filtered string-based search option was used with the following user-defined search parameters:

Factor Selection Attribute: Organism Classification
Search Pattern: Mammalia
Max. Allowable Mismatch %: 0
Min. element length: 5
Min. log-likelihood: 10

This parameter selection specifies that only mammalian TF binding sites (approximately 1,400 of the 4,401 entries in the database) that are at least 5 bases long will be included in the search. It further specifies that only TF binding sites that have a perfect match in the query sequence and a minimum log likelihood (LLH) score of 10 will be reported. The LLH scoring method assigns 2 to an unambiguous match, 1 to a partially ambiguous match (e.g., A or T match W) and 0 to a match against ‘N’. For example, a search with parameters specified above would result in a “hit” (positive result or match) for TATAAA (SEQ ID NO:240) (LLH=10), STRATG (SEQ ID NO:241) (LLH=10), and MTNCGNMA (SEQ ID NO:242) (LLH=10) but not for TRATG (SEQ ID NO: 243) (LLH=9) if these four TF binding sites were present in the query sequence. A lower stringency test was performed at the end of the design process to re-evaluate the search parameters.

When TESS was tested with a mock query sequence containing known TF binding sites it was found that the program was unable to report matches to sites ending with the 3’ end of the query sequence. Thus, an extra nucleotide was added to the 3’ end of all query sequences to eliminate this problem.

The first search for TF binding sites using the parameters described above found about 100 transcription factor binding sites (hits) for each of the two synthetic genes (GRver2 and RDver2). All sites were eliminated by changing one or more codons of the synthetic gene sequences in accordance with the codon optimization guidelines described in 1a above. However, it was expected that some of these changes created new TF binding sites, other regulatory sites, and new restriction sites. Thus, steps 2 a-d were repeated as described, and 4 new restriction sites and 2 new splice sites were removed. The two output sequences from this third design step were named GRver3 and RDver3. Their DNA sequences are 66% identical (541 mismatches) (FIGS. 2 and 3).

4. Remove New Transcription Factor (TF) Binding Sites, then Repeat Steps 2 a-d

The starting gene sequences for this design step were GRver3 and RDver3. This fourth step is an iteration of the process described in step 3. The search for newly introduced TF binding sites yielded about 50 hits for each of the two synthetic genes. All sites were eliminated by changing one or more codons of the synthetic gene sequences in general accordance with the codon optimization guidelines described in 1a above. However, more high to medium usage codons were used to allow elimination of all TF binding sites. The lowest priority was placed on maintaining low sequence identity between the GR and RD genes. Then steps 2 a-d were repeated as described. The two output sequences from this fourth design step were named GRver4
and RDver4. Their DNA sequences are 68% identical (506 mismatches) (FIGS. 2 and 3).

5. Remove New Transcription Factor (TF) Binding Sites, then Repeat Steps 2-4.

[0172] The starting gene sequences for this design step were GRver4 and RDver4. This fifth step is another iteration of the process described in step 3 above. The search for new TF binding sites introduced in step 4 yielded about 20 hits for each of the two synthetic genes. All sites were eliminated by changing one or more codons of the synthetic gene sequences in general accordance with the codon optimization guidelines described in 1a above. However, more high to medium usage codons were used (these are all considered “preferred”) to allow elimination of all TF binding sites. The lowest priority was placed on maintaining low sequence identity between the GR and RD genes. Then steps 2-4 were repeated as described. Only one acceptor splice site could not be eliminated. As a final step the absence of all TF binding sites in both genes as specified in step 3 was confirmed. The two output sequences from this fifth and last design step were named GRver5 and RDver5. Their DNA sequences are 69% identical (504 mismatches) (FIGS. 2 and 3).

Additional Evaluation of GRver5 and RDver5:

a) Use Lower Stringency Parameters for TESS:

[0173] The search for TF binding sites was repeated as described in step 3 above, but with even less stringent user-defined parameters:

[0174] setting LLH to 9 instead of 10 did not result in new hits;

[0175] setting LLH to 0 through 8 (incl.) resulted in hits for two additional sites, MAMAG (22 hits) and CTKKK (24 hits);

[0176] setting LLH to 8 and the minimum element length to 4, the search yielded (in addition to the two sites above) different 4-base sites for AP-1, NF-1, and c-Myb that are shortened versions of their longer respective consensus sites which were eliminated in steps 3-5 above.

It was not realistic to attempt complete elimination of these sites without introduction of new sites, so no further changes were made.

b) Search Different Database:

[0177] The Eukaryotic Promoter Database (release 45) contains information about reliably mapped transcription start sites (1253 sequences) of eukaryotic genes. This database was searched using BLASTN 1.4.11 with default parameters (optimized to find nearly identical sequences rapidly; see Altschul et al., 1990) at the National Center for Biotechnology Information site (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST). To test this approach, a portion of pG3-Control vector sequence containing the SV40 promoter and enhancer was used as a query sequence, yielding the expected hits to SV40 sequences. No hits were found when using the two synthetic genes as query sequences.

Summary of GRver5 and RDver5 Synthetic Gene Properties:

[0178] Both genes, which at this stage were still only “virtual” sequences in the computer, have a codon usage that strongly favors mammalian high-usage codons and minimizes mammalian and E. coli low-usage codons. FIG. 4 shows a summary of the codon usage of the parent gene and the various synthetic gene versions.

[0179] Both genes are also completely devoid of eukaryotic TF binding sites consisting of more than four unambiguous bases, donor and acceptor splice sites (one exception: GRver5 contains one splice acceptor site), poly(A) addition sites, specific prokaryotic (E. coli) regulatory sequences, and undesired restriction sites.

[0180] The gene sequence identity between GRver5 and RDver5 is only 69% (504 base mismatches) while their encoded proteins are 99% identical (4 amino acid mismatches), see FIGS. 2 and 3. Their identity with the parent sequence YG81-6G1 is 74% (GRver5) and 73% (RDver5), see FIG. 2. Their base composition is 49.9% GC (GRver5) and 49.5% GC (RDver5), compared to 40.2% GC for the parent YG81-6G01.

Construction of Synthetic Genes:

[0181] The two synthetic genes were constructed by assembly from synthetic oligonucleotides in a thermocycler followed by PCR amplification of the full-length genes (similar to Stemmer et al. (1995) Gene, 164, pp. 49-53). Unintended mutations that interfered with the design goals of the synthetic genes were corrected.

a) Design of Synthetic Oligonucleotides:

[0182] The synthetic oligonucleotides were mostly 40mers that collectively code for both complete strands of each designed gene (1,626 bp) plus flanking regions needed for cloning (1,950 bp total for each gene; FIG. 6). The 5' and 3' boundaries of all oligonucleotides specifying one strand were generally placed in a manner to give an average offset/overlap of 20 bases relative to the boundaries of the oligonucleotides specifying the opposite strand.

[0183] The ends of the flanking regions of both genes matched the ends of the amplification primers (pRAMtailp: 5'gtcatcagctgtagcgcgccgccagctggtcatcagctgtagcgcgccgccagcagcgcgccgccagcagcgcgccgcaagcgcgag SEQ ID NO:229, and pRAMtaild: 5'catcagctgtagcgcgccgccagctggtcatcagctgtagcgcgccgccagcagcgcgccgccagcagcgcgccgcaagcgcgag SEQ ID NO:230) to allow cloning of the genes into our E. coli expression vector pRAM (WO99/14336).

[0184] A total of 183 oligonucleotides were designed (FIG. 6): fifteen oligonucleotides that collectively encode the upstream and downstream flanking sequences (identical for both genes; SEQ ID NOs: 35-49) and 168 oligonucleotides (4×42) that encode both strands of the two genes (SEQ ID NOs: 50-217).

[0185] All 183 oligonucleotides were run through the hairpin analysis of the Oligo software (Oligo 4.0 Primer Analysis Software © 1989-1991 by Wojciech Rychlik) to identify potentially detrimental intra-molecular loop formation. The guidelines for evaluating the analysis results were set according to recommendations of Dr. Sims (Sigma-Genosys Custom Gene Synthesis Department): oligos forming hairpins with ΔG< -10 have to be avoided, those forming hairpins with ΔG between -7 and -10 involving the 3' end of the oligonucleotide should also be avoided, while those with an overall ΔG between -5 should not pose a problem for this application. The analysis identified 23 oligonucleotides able to form hairpins with a ΔG between -7.1 and -4.9. Of these, 5 had blocked
or nearly blocked 3' ends (0-3 free bases) and were redesigned by removing 1-4 bases at their 3' end and adding it to the adjacent oligonucleotide.

[0186] The 40mer oligonucleotide covering the sequence complementary to the poly(A) tail had a very low complexity 3' end (13 consecutive T bases). An additional 40mer was designed with a high complexity 3' end but a consequently reduced overlap with one of its complementary oligonucleotides (11 instead of 20 bases) on the opposite strand.

[0187] Even though the oligos were designed for use in a thermocycler-based assembly reaction, they could also be used in a ligation-based protocol for gene construction. In this approach, the oligonucleotides are annealed in a pairwise fashion and the resulting short double-stranded fragments are ligated using the sticky overhangs. However, this would require that all oligonucleotides be phosphorylated.

b) Gene Assembly and Amplification

[0188] In a first step, each of the two synthetic genes was assembled in a separate reaction from 98 oligonucleotides. The total volume for each reaction was 50 μl:

[0189] 0.5 μM oligonucleotides (≈0.25 pmoles of each oligo)
[0190] 1.0 U Taq DNA polymerase
[0191] 0.02 U Pfu DNA polymerase
[0192] 2 mM MgCl₂
[0193] 0.2 mM dNTPs (each)
[0194] 0.1% gelatin
[0195] Cycling conditions: (94°C for 30 seconds, 52°C for 30 seconds, and 72°C for 30 seconds)×5 cycles.

[0196] In a second step, each assembled synthetic gene was amplified in a separate reaction. The total volume for each reaction was 50 μl:

[0197] 2.5 μl assembly reaction
[0198] 5.0 U Taq DNA polymerase
[0199] 0.1 U Pfu DNA polymerase
[0200] 1 M each primer (pRAMtailup, pRAMtailldn)
[0201] 2 mM MgCl₂
[0202] 0.2 mM dNTPs (each)
[0203] Cycling conditions: (94°C for 20 seconds, 65°C for 60 seconds, 72°C for 3 minutes)×30 cycles.

[0204] The assembled and amplified genes were subcloned into the pRAM vector and expressed in E. coli, yielding 75% luminescent GR or RD clones. Forty-four GR and 44 RD clones were analyzed with our screening robot (WO99/14336). The six best GR and RD clones were manually analyzed and one best GR and RD clone was selected (GR6 and RD7). Sequence analysis of GR6 revealed two point mutations in the coding region, both of which resulted in an amino acid substitution (G49N and P230S). Sequence analysis of RD7 revealed three point mutations in the coding region, one of which resulted in an amino acid substitution (H36Y). It was confirmed that none of the silent point mutations introduced any regulatory or restriction sites conflicting with the overall design criteria for the synthetic genes.

c) Corrective Assembly and Amplification

[0205] To remove the large number of mutations present in the full-length synthetic genes we performed an additional assembly and amplification reaction for each gene using the proof-reading DNA polymerase Tli. The assembly reaction contained, in addition to the 98 GR or RD oligonucleotides, a small amount of DNA from the corresponding full-length clones with mutations described above. This allows the oligos to correct mutations present in the templates.

[0206] The following assembly reaction was performed for each of the synthetic genes. The total volume for each reaction was 50 μl:

[0207] 0.5 μM oligonucleotides (≈0.25 pmoles of each oligo)
[0208] 0.016 pmol plasmid (mix of clones with correct insert size)
[0209] 2.5 U Tli DNA polymerase
[0210] 2 mM MgCl₂
[0211] 0.2 mM dNTPs (each)
[0212] 0.1% gelatin
[0213] Cycling conditions: 94°C for 30 seconds, then (94°C for 30 seconds, 52°C for 30 seconds, and 72°C for 30 seconds) for 55 cycles.

[0214] The following amplification reaction was performed on each of the assembly reactions. The total volume for each amplification reaction was 50 μl:

[0215] 1-5 μl of assembly reaction
[0216] 40 pmol each primer (pRAMtailup, pRAMtailldn)
[0217] 2.5 U Tli DNA polymerase
[0218] 2 mM MgCl₂
[0219] 0.2 mM dNTPs (each)
[0220] Cycling conditions: 94°C for 30 seconds, then (94°C for 20 seconds, 65°C for 60 seconds and 72°C for 5 minutes) for 30 cycles.

[0221] The genes obtained from the corrective assembly and amplification step were subcloned into the pRAM vector and expressed in E. coli, yielding 75% luminescent GR or RD clones. Forty-four GR and 44 RD clones were analyzed with our screening robot (WO99/14336). The six best GR and RD clones were manually analyzed and one best GR and RD clone was selected (GR6 and RD7). Sequence analysis of GR6 revealed two point mutations in the coding region, both of which resulted in an amino acid substitution (G49N and P230S). Sequence analysis of RD7 revealed three point mutations in the coding region, one of which resulted in an amino acid substitution (H36Y). It was confirmed that none of the silent point mutations introduced any regulatory or restriction sites conflicting with the overall design criteria for the synthetic genes.

d) Reversal of Untended Amino Acid Substitutions

[0222] The unintended amino acid substitutions present in the GR6 and RD7 synthetic genes were reversed by site-directed mutagenesis to match the GRver5 and RDver5 designed sequences, thereby creating GRver5.1 and
RDver5.1. The DNA sequences of the mutated regions were confirmed by sequence analysis.

c) Improve Spectral Properties

[0223] The RDver5.1 gene was further modified to improve its spectral properties by introducing an amino change (R351G), thereby creating RDver5.2

pGL3 Vectors with RD and GR Genes

[0224] The parent click beetle luciferase YGI#81-6G1 (“YG”), and the synthetic click beetle luciferase genes GRver5.1 (“GR”), RDver5.2 (“RD”), and RD156-1H9 were cloned into the four pGL3 reporter vectors (Promega Corp.):

[0225] pGL3-Basic—no promoter, no enhancer
[0226] pGL3-Control—SV40 promoter, SV40 enhancer
[0227] pGL3-Enhancer—SV40 enhancer (3’ to luciferase coding sequences)
[0228] pGL3-Promoter—SV40 promoter.

[0229] The primers employed in the assembly of GR and RD synthetic genes facilitated the cloning of those genes into pRAM vectors. To introduce the genes into pGL3 vectors (Promega Corp., Madison, Wis.) for analysis in mammalian cells, each gene in a pRAM vector (pRAM RDver5.1, pRAM GRver5.1, and pRAM RD156-1H9) was amplified to introduce an Nco I site at the 5’ end and an Xba I site at the 3’ end of the gene. The primers for pRAM RDver5.1 and pRAM GRver5.1 were:

GR—5’ GGA TCC CAT GGT GAA GCG TGA GAA 3’

or

[SEQ ID NO:231]

(SEQ ID NO:231)
GR—5’ GGA TCC CAT GGT GAA GCG TGA GAA 3’

[SEQ ID NO:232]

RD—5’ GGA TCC CAT GGT GAA ACG CGA 3’

and

[SEQ ID NO:233]

5’ CTA GCT TTG TTT TCT AGA TAA TGA AGA C 3’

[0230] The primers for pRAM RD156-1H9 were:

5’ GCG TAT CCA TGG TAA AGC GTC AGA AAA ATG TG 3’

and

[SEQ ID NO:295]

(SEQ ID NO:295)
5’ GCG TAT CCA TGG TAA AGC GTC AGA AAA ATG TG 3’

(SEQ ID NO:296)

5’ CCG ACT CTA GAT TAC TAA CGG GCC GCC TTC ACC 3’

[0230] The PCR included:

[0231] 100 ng DNA plasmid
[0232] 1 µM primer upstream
[0233] 1 µM primer downstream
[0234] 0.2 mM dNTPs
[0235] 1x buffer (Promega Corp.)
[0236] 5 units Pfu DNA polymerase (Promega Corp.)
[0237] Sterile nanopure H2O to 50 µl

[0238] The cycling parameters were: 94° C. for 5 minutes; (94° C. for 30 seconds, 55° C. for 1 minute; and 72° C. for 3 minutes) x15 cycles. The purified PCR product was digested with Neo I and Xba I, ligated with pGL3-control that was also digested with Neo I and Xba I, and the ligated products introduced to E. coli. To insert the luciferase genes into the other pGL3 reporter vectors (basic, promoter and enhancer), the pGL3-control vectors containing each of the luciferase genes was digested with Neo I and Xba I, ligated with other pGL3 vectors that also were digested with Neo I and Xba I, and the ligated products introduced to E. coli. Note that the polypeptide encoded by GRver5.1 and RDver5.1 (and RD 156-1 H9, see below) nucleic acid sequences in pGL3 vectors has an amino acid substitution at position 2 to valine as a result of the Neo I site at the initiation codon in the oligonucleotide.

[0239] Because of internal Neo I and Xba I sites, the native gene in YGI #81-6G01 was amplified from a Hind III site upstream to a Hpa I site downstream of the coding region and which included flanking sequences found in the GR and RD clones. The upstream primer (5’-CAA AAA GCT TGG CAT TCC GGT ACT GTT GGT AAA GCC ACC ATG GTG AAG CGA GAG-3’, SEQ ID NO:234) and a downstream primer (5’-CAA TTG TTT TGG TTA ACC TGT TTA TIT GC-3’, SEQ ID NO:235) were mixed with YGI#81-6G01 and amplified using the PCR conditions above. The purified PCR product was digested with Neo I and Xba I, ligated with pGL3-control that was also digested with Hind III and Hpa I, and the ligated products introduced into E. coli. To insert YGI#81-6G01 into the other pGL3 reporter vectors (basic, promoter and enhancer), the pGL3-control vectors containing YGI#81-6G01 were digested with Neo I and Xba I, ligated with the other pGL3 vectors that also were digested with Neo I and Xba I, and the ligated products introduced to E. coli. Note that the clone of YGI#81-6G01 in the pGL3 vectors has a C instead of an A at base 786, which yields a change in the amino acid sequence at residue 262 from Phe to Leu (FIG. 2 shows the sequence of YGI#81-6G01 prior to introduction into pGL3 vectors). To determine whether the altered amino acid at position 262 affected the enzyme biochemistry, the clone of YGI#81-6G01 was mutated to resemble the original sequence. Both clones were then tested for expression in E. coli, physical stability, substrate binding, and luminescence output kinetics. No significant differences were found.

[0240] Partially purified enzymes expressed from the synthetic genes and the parent gene were employed to determine Km for luciferin and ATP (see Table 3).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Km (LH2)</th>
<th>Km (ATP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YG parent</td>
<td>2 µM</td>
<td>17 µM</td>
</tr>
<tr>
<td>GR</td>
<td>1.3 µM</td>
<td>25 µM</td>
</tr>
<tr>
<td>RD</td>
<td>26.5 µM</td>
<td>46 µM</td>
</tr>
</tbody>
</table>

[0241] In vitro eukaryotic transcription/translation reactions were also conducted using Promega’s TNT T7 Quick system according to manufacturer’s instructions. Luminescence levels were 1 to 37-fold and 1 to 77-fold higher (depending on the reaction time) for the synthetic GR and RD genes, respectively, compared to the parent gene (corrected for luminometer spectral sensitivity).

[0242] To test whether the synthetic click beetle luciferase genes and the wild type click beetle gene have improved
expression in mammalian cells, each of the synthetic genes and the parent gene was cloned into a series of pGL3 vectors and introduced into CHO cells (Table 8). In all cases, the synthetic click beetle genes exhibited a higher expression than the native gene. Specifically, expression of the synthetic GR and RD genes was 1900-fold and 40-fold higher, respectively, than that of the parent (transfection efficiency normalized by comparison to native Renilla luciferase gene). Moreover, the data (basic versus control vector) show that the synthetic genes have reduced basal level transcription.

[0243] Further, in experiments with the enhancer vector where the percentage of activity in reference to the control is compared between the native and synthetic gene, the data showed that the synthetic genes have reduced risk of anomalous transcription characteristics. In particular, the parent gene appeared to contain one or more internal transcriptional regulatory sequences that are activated by the enhancer in the vector, and thus is hot suitable as a reporter gene while the synthetic GR and RD genes showed a clean reporter response (transfection efficiency normalized by comparison to native Renilla luciferase gene). See Table 9.

[0244] The clone names and their corresponding SEQ ID numbers for nucleotide sequence and amino acid sequence are listed below in Table 4.

<table>
<thead>
<tr>
<th>Clone name</th>
<th>Luciferase Type</th>
<th>SEQ ID NO.</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUC/PL3YG</td>
<td>Wild type YG Click Beetle</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>YG(81-60G01)</td>
<td>Mutant YG Click Beetle</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>GVer1</td>
<td>Synthetic Green Click Beetle</td>
<td>3</td>
<td>25</td>
</tr>
<tr>
<td>GVer2</td>
<td>Synthetic Green Click Beetle</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>GVer3</td>
<td>Synthetic Green Click Beetle</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>GVer4</td>
<td>Synthetic Green Click Beetle</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>GVer5</td>
<td>Synthetic Green Click Beetle</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>GVer6</td>
<td>Synthetic Green Click Beetle</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>GVer5.1</td>
<td>Synthetic Green Click Beetle</td>
<td>9</td>
<td>31</td>
</tr>
<tr>
<td>RDVer1</td>
<td>Synthetic Red Click Beetle</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>RDVer2</td>
<td>Synthetic Red Click Beetle</td>
<td>11</td>
<td>33</td>
</tr>
<tr>
<td>RDVer3</td>
<td>Synthetic Red Click Beetle</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>RDVer4</td>
<td>Synthetic Red Click Beetle</td>
<td>13</td>
<td>35</td>
</tr>
<tr>
<td>RDVer5</td>
<td>Synthetic Red Click Beetle</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>RDVer6</td>
<td>Synthetic Red Click Beetle</td>
<td>15</td>
<td>37</td>
</tr>
<tr>
<td>RDVer5.1</td>
<td>Synthetic Red Click Beetle</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>RDVer5.2</td>
<td>Synthetic Red Click Beetle</td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>RD156-H119</td>
<td>Synthetic Red Click Beetle</td>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td>REIUC</td>
<td>Wild type Renilla</td>
<td>19</td>
<td>224</td>
</tr>
<tr>
<td>RUCver1</td>
<td>Synthetic Renilla</td>
<td>20</td>
<td>225</td>
</tr>
<tr>
<td>RUCver2</td>
<td>Synthetic Renilla</td>
<td>21</td>
<td>226</td>
</tr>
<tr>
<td>RUC-fin1</td>
<td>Synthetic Renilla</td>
<td>22</td>
<td>227</td>
</tr>
</tbody>
</table>

**TABLE 4**

**EXAMPLE 2**

Evolution of the RD Luciferase Gene

[0245] RDVer5.2 was mutated to increase its luminescence intensity, thereby creating RD156-H19 which carries four additional amino acid changes (M21, S349T, K488T, E538V) and three silent point mutations (SEQ ID NO:18).

a) Site-Directed Mutagenesis:

[0246] The initial strategy was to use site-directed mutagenesis. There are four amino acid differences between the GR and RD synthetic genes with I348Q providing the greatest contribution to red color. Thus, this substitution may also cause structural changes in the protein that could lead to low light output. Optimization of positions near this area could increase light output. The following positions were selected for mutagenesis:

- [0247] 1. S344 (at the edge of the binding pocket for luciferin)—randomize this codon.
- [0248] 2. A245 (strictly conserved but closest to 348 and at the edge of the active site pocket)—randomize this codon.
- [0249] 3. L347 (not conserved, next to 348 in sequence)—mutate to hydrophobic amino acids only.
- [0250] 4. S349 (not conserved, next to 348 in sequence)—mutate to S, T, A, P only.

[0251] Oligonucleotides designed to mutate the above positions were used in a site-directed mutagenesis experiment (WO99/14336) and the resulting mutants were screened for luminescence intensity. There was little variation in light intensity and only about 25% were luminescent. For more detailed analysis, clones were picked and analyzed with the screening robot (PCT/WO9914336). None of the clones had a luminescence intensity (LI) higher than RDver5.2, but four of the clones had slightly lower composite Km for luciferin and ATP (Km).

b) Directed Evolution:

[0252] Protocols and procedures used for the directed evolution are detailed in see PCT/WO9914336. DNA from the four clones with lower Km was combined and three libraries of random mutants were produced. The libraries were screened with the robot and clones with the highest LI values were selected. These clones were shuffled together and another robotic screen was completed with an incubation temperature of 46°C. The three clones with the highest LI values were RD156-0B4, RD156-1A5, and RD156-119.

c) Analysis:

[0253] The three clones with the highest LI values were selected for manual analysis to confirm that their luminescence intensity was higher than that of RDVer5.2 and to ensure that their spectral properties were not compromised. One of the clones was slightly green-shifted, all others maintained the spectral properties of RDVer5.2 (Table 5).

<table>
<thead>
<tr>
<th>Clone</th>
<th>Peak (nm)</th>
<th>Width (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD156-0B4</td>
<td>616</td>
<td>68</td>
</tr>
<tr>
<td>RD156-1A5</td>
<td>614</td>
<td>70</td>
</tr>
<tr>
<td>RD156-119</td>
<td>618</td>
<td>69</td>
</tr>
<tr>
<td>RDVer5.2 (prep #1)</td>
<td>617</td>
<td>70</td>
</tr>
<tr>
<td>RDVer5.2 (prep #2)</td>
<td>618</td>
<td>69</td>
</tr>
</tbody>
</table>

[0254] The Km values for luciferin and the luminescence intensity relative to RDVer5.2 were determined for all three clones in several independent experiments. All cells samples were processed with CCLR lysis buffer (E1483, Promega Corp., Madison, Wis.) and diluted 1:10 into buffer (25 mM HEPES pH 7.8, 5% glycerol, 1 mg/ml BSA, 150 mM NaCl). Table 7 summarizes the results (Lumen: luminescence values were normalized to optical density; measurements for independent experiments are separated by forward slashes) from expression in bacterial cells. RD156-119, the clone with the
highest luminescence intensity (5 to 10-fold increase) also has an about 2-fold higher Km for luciferin.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Km Luciferin [µM]</th>
<th>Lum (normalized to RDver5.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD156-1H9</td>
<td>8/10</td>
<td>2.2/2.5</td>
</tr>
<tr>
<td>RD156-1A5</td>
<td>13/13</td>
<td>3.1/3.5</td>
</tr>
<tr>
<td>RD156-1H9</td>
<td>20/23/23</td>
<td>4/10.9/7.5</td>
</tr>
<tr>
<td>RDver5.2 (prep #1)</td>
<td>12/14/14</td>
<td></td>
</tr>
<tr>
<td>RDver5.2 (prep #2)</td>
<td>40/50</td>
<td></td>
</tr>
<tr>
<td>GRver5.1 (prep #1)</td>
<td>0.5</td>
<td>64</td>
</tr>
<tr>
<td>GRver5.1 (prep #2)</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 7 shows a comparison between the luminescence intensities of RD156-1H9, GRver5.1 and RDver5.2 normalized to GRver5.1 with and without correction for the spectral sensitivity of the luminometer photomultiplier tube. With correction, the luminescence intensity of clone RD156-1H9 was only about 2-fold lower than that of GRver5.1. The luciferin Km for clone RD156-1H9 is approximately 40-fold higher than GRver5.1. RD156-1H9 is thermostable at 50°C, for at least 2 hours.

Table 8 and 9 show a comparison of luciferase expression levels in CHO cells. Table 8 shows the expression levels only from the control vectors in comparison to the firefly luciferase gene (RLU-relative light units). Table 9 shows a comparison of the expression levels in all four pGL3 vectors calculated as a percent of the expression level in pGL3-control.

**EXAMPLE 3**

**Synthetic Renilla Luciferase Nucleic Acid Molecule**

The synthetic Renilla luciferase genes prepared include 1) an introduced Kozak sequence, 2) codon usage optimized for mammalian (human) expression, 3) a reduction or elimination of unwanted restriction sites, 4) removal of prokaryotic regulatory sites (ribosome binding site and TATA box), 5) removal of splice sites and poly(A) addition sites, and 6) a reduction or elimination of mammalian transcriptional factor binding sequences.

The process of computer-assisted design of synthetic Renilla luciferase genes by iterative rounds of codon optimization and removal of transcription factor binding sites and other regulatory sites as well as restriction sites can be described in three steps:

1. Using the wild type Renilla luciferase gene as the parent gene, codon usage was optimized, one amino acid was changed (T→A) to generate a Kozak consensus sequence, and undesired restriction sites were eliminated thereby creating synthetic gene Rhuever1.
2. Remove prokaryotic regulatory sites, splice sites, poly(A) sites and transcription factor (TF) binding sites (first pass). Then remove newly created TF binding sites. Then remove newly created undesired restriction enzyme sites, prokaryotic regulatory sites, splice sites, and poly(A) sites without introducing new TF binding sites. This thereby created Rhuever2.

4. The actual gene was then constructed from synthetic oligonucleotides corresponding to the Rhuc-final designed sequence. All mutations resulting from the assembly or PCR process were corrected. This gene is Rhuc-final (SEQ ID NO:22) and encodes the amino acid sequence of SEQ ID NO:227.

**Codon Selection**

Starting with the Renilla reniformis luciferase sequence in Genbank (Accession No. M63501, SEQ ID NO: 19), codons were selected based on codon usage for optimal expression in human cells and to avoid E. coli low-usage codons. The best codon for expression in human cells (or the
best two codons if found at a similar frequency) was chosen for all amino acids with more than one codon (Wada et al., 1990):

- Arg: CGC
- Lys: AAG
- Leu: CUG
- Ser: UUC/AGC
- Thr: ACC
- Pro: CCA/CCT
- Ala: GCC
- Gly: GCC
- Val: GTG
- Ile: ATC/ATT
- Phe: TTC

In cases where two codons were selected for one amino acid, they were used in an alternating fashion. To meet other criteria for the synthetic gene, the initial optimal codon selection was modified to some extent later. For example, introduction of a Kozak sequence required the use of GCT for Ala at amino acid position 2 (see below).

The following low-usage codons in mammalian cells were not used unless needed: Arg: CGA, CGU; Leu: CTA, UUA; Ser: TCG; Pro: CCG; Val: GTA; and Ile: AIA. The following low-usage codons in E. coli were also avoided when reasonable (note that 3 of these match the low-usage list for mammalian cells): Arg: CGA/CGG/AGA/AGG, Leu: CTA; Pro: CCG; Ile: AIA.

Introduction of Kozak Sequences

The Kozak sequence: 5’ aaccATGGCT 3’ (SEQ ID NO: 293) (the Nco 1 site is underlined, the coding region is shown in capital letters) was introduced to the synthetic Renilla luciferase gene. The introduction of the Kozak sequence changes the second amino acid from Thr to Ala (GCT).

Removal of Undesired Restriction Sites

REBASE was employed to identify undesirable restriction sites as described in Example 1. The following undesired restriction sites (in addition to those described in Example 1) were removed according to the process described in Example 1: EcoCR I, NdeI, NsiI, SphI, Spel, Xmal, PstI.

The version of Renilla luciferase (Rhu) which incorporates all these changes is Rhuver1.

Removal of Prokaryotic (E. coli) Regulatory Sequences Splice Sites, and Poly(A) Sites

The priority and process for eliminating transcription regulation sites was as described in Example 1.

Removal of TF Binding Sites

The same process, tools, and criteria were used as described in Example 1, however, the newer version 3.3 of the TRANSFAC database was employed.

After removing prokaryotic regulatory sequences, splice sites and poly(A) sites from Rluver1, the first search for TF binding sites identified about 60 hits. All sites were eliminated with the exception of three that could not be removed without altering the amino acid sequence of the synthetic Renilla gene:

1. site at position 63 composed of two codons for W (TGGTTG), for CAC-binding protein T00076;
2. site at position 522 composed of codons for KMV (AAN ATG GTN), for myc-DF1 T00517;
3. site at position 885 composed of codons for EMG (GAR ATG GGN), for myc-DF1 T00517.

The subsequent second search for (newly introduced) TF binding sites yielded about 20 hits. All new sites were eliminated, leaving only the three sites described above. Finally, any newly introduced restriction sites, prokaryotic regulatory sequences, splice sites and poly(A) sites were removed without introducing new TF binding sites if possible.

Rluver2 was obtained (SEQ ID Nos. 21 and 226).

As in Example 1, lower stringency search parameters were specified for the TESS filtered string search to further evaluate the synthetic Renilla gene.

With the LLH reduced from 10 to 9 and the minimum element length reduced from 5 to 4, the TESS filtered string search did not show any new hits. When, in addition to the parameter changes listed above, the organism classification was expanded from “mammal” to “chordata”, the search yielded only four more TF binding sites. When the Min LLH was further reduced to between 8 and 0, the search showed two additional 5-base sites (MAMAG and CTKTK) which combined had four matches in Rluver2, as well as several 4-base sites. Also as in Example 1, Rluver2 was checked for hits to entries in the EFD (Eukaryotic Promoter Database, Release 45). Three hits were determined (one to Mus musculus promoter H-2L’d (Cell, 44, 261 (1986), one to Herpes Simplex Virus type 1 promoter b’g2.7 kb, and one to Homo sapiens DHFR promoter (J. Mol. Biol., 176, 169 (1984)). However, no further changes were made to Rluver2.

Summary of Properties for Rluver2

All 30 low usage codons were eliminated. The introduction of a Kozak sequence changed the second amino acid from Thr to Ala;

base composition: 55.7% GC (Renilla wild-type parent gene: 36.5%);

one undesired restriction site could not be eliminated: EcoR V at position 488;

the synthetic gene had no prokaryotic promoter sequence but one potentially functional ribosome binding site (RBS) at positions 867-73 (about 13 bases upstream of a Met codon) could not be eliminated;

all poly(A) addition sites were eliminated;

splice sites: 2 donor splice sites could not be eliminated (both share the amino acid sequence MGK);

TF sites: all sites with a consensus of 4 unambiguous bases were eliminated (about 280 TF binding
sites were removed) with 3 exceptions due to the preference to avoid changes to the amino acid sequence.

Synthetic Renilla luciferase sequences are shown in FIGS. 7 and 8. A codon usage comparison is shown in FIG. 9.

[0286] When introduced into pGL3, RLuc-final has a Kozak sequence (CACCATGGCT). The changes in RLuc-final relative to RLucer2 were introduced during gene assembly. One change was at position 619, a C to an A, which eliminated a eukaryotic promoter sequence and reduced the stability of a hairpin structure in the corresponding oligonucleotide employed to assemble the gene. Other changes included a change from CGC to AGA at positions 218-220 (resulted in a better oligonucleotide for PCR).

Gene Assembly Strategy

[0287] The gene assembly protocol employed for the synthetic Renilla luciferase was similar to that described in Example 1. The oligonucleotides employed are shown in FIG. 10.

Sense Strand primer: 5’ AACCAGGGCCCTCAGGCTTACGACCCCAGCAGGCAAAGTA 3’

Anti-sense Strand primer: 5’ GCTCTAGATTAATTGGCCCTTATTTATAGGCGCTTCCAG 3’

[0288] The resulting synthetic gene fragment was cloned into a pRAM vector using Neo I and Xba I. Two clones having the correct size insert were sequenced. Four to six mutations were found in the synthetic gene from each clone. These mutations were fixed by site-directed mutagenesis (Gene Editor from Promega Corp., Madison, Wis.) and swapping the correct regions between these two genes. The corrected gene was confirmed by sequencing.

Other Vectors

[0289] To prepare an expression vector for the synthetic Renilla luciferase gene in a pGL-3 control vector backbone, 5 μg of pGL-3-control was digested with Neo I and Xba I in 50 μl final volume with 2 μl of each enzyme and 5 μl 10x buffer B (nuclease water was used to fill the volume to 50 μl). The digestion reaction was incubated at 37°C for 2 hours, and the whole mixture was run on a 1% agarose gel in 1xTAE. The desired vector backbone fragment was purified using Qiagen’s QIAquick gel extraction kit.

[0290] The native Renilla luciferase gene fragment was cloned into pGL3-control vector using two oligonucleotides, Neo I-RL-F and Xba I-RL-R, to PCR amplify native Renilla luciferase gene using pRL-CMV as the template. The sequence for Neo I-RL-F is 5’-CGCTAGCCATGGCTTACGACCCCAGGCAAAGTATAATC’-3’ (SEQ ID NO:238); the sequence for Xba I-RL-R is 5’-GCCCAGGTACACTCTAGAATATTGTGT-3’ (SEQ ID NO:239). The PCR reaction was carried out as follows:

[0291] Reaction mixture (for 100 μl):

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA template (Plasmid)</td>
<td>1.0 μl (1.0 ng/μl final)</td>
</tr>
<tr>
<td>10x Rec. Buffer</td>
<td>10.0 μl (Stratagene Corp.)</td>
</tr>
<tr>
<td>dNTPs (25 mM each)</td>
<td>1.0 μl (final 250 μM)</td>
</tr>
<tr>
<td>Primer 1 (10 μM)</td>
<td>2.0 μl (0.2 μM final)</td>
</tr>
<tr>
<td>Primer 2 (10 μM)</td>
<td>2.0 μl (0.2 μM final)</td>
</tr>
<tr>
<td>Phi DNA Polymerase</td>
<td>2.0 μl (2.5 U/μl, Stratagene Corp.)</td>
</tr>
<tr>
<td>82.0 μl double distilled water</td>
<td></td>
</tr>
</tbody>
</table>

[0292] PCR Reaction: heat 94°C for 2 minutes; (94°C for 20 seconds; 65°C for 1 minute; 72°C for 2 minutes; then 72°C for 5 minutes) x 25 cycles, then incubate on ice. The PCR amplified fragment was cut from a gel, and the DNA purified and stored at −20°C.

[0293] To introduce native Renilla luciferase gene fragment into pGL3-control vector, 5 μg of the PCR product of the native Renilla luciferase gene (RAM-RL-synthetic) was digested with Neo I and Xba I. The desired Renilla luciferase gene fragment was purified and stored at −20°C.

[0294] Then 100 ng of insert and 100 ng of pGL3-control vector backbone were digested with restriction enzymes Neo I and Xba I and ligated together. Then 2 μl of the ligation mixture was transformed into JM109 competent cells. Eight ampicillin resistance clones were picked and their DNA isolated. DNA from each positive clone of pGL3-control-native and pGL3-control-synthetic was purified. The correct sequences for the native gene and the synthetic gene in the vectors were confirmed by DNA sequencing.

[0295] To determine whether the synthetic Renilla luciferase gene has improved expression in mammalian cells, the gene was cloned into the mammalian expression vector pGL3-control vector under the control of SV40 promoter and SV40 early enhancer (FIG. 13A). The native Renilla luciferase gene was also cloned into the pGL3-control vector so that the expression from synthetic gene and the native gene could be compared. The expression vectors were then transfected into four common mammalian cell lines (CHO, NIH3T3, Hela and CV-1; Table 10), and the expression levels compared between the vectors with the synthetic gene versus the native gene. The amount of DNA used was at two different levels to ascertain that expression from the synthetic gene is consistently increased at different expression levels. The results show a 70-600 fold increase of expression for the synthetic Renilla luciferase gene in these cells (Table 10).

<table>
<thead>
<tr>
<th>Enhanced Synthetic Renilla Gene Expression</th>
<th>Cell Type</th>
<th>Amount Vector</th>
<th>Fold Expression Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO</td>
<td>0.2 μg</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.8 μg</td>
<td>145</td>
<td></td>
</tr>
<tr>
<td>NIH3T3</td>
<td>0.2 μg</td>
<td>326</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0 μg</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>HeLa</td>
<td>0.2 μg</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0 μg</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>CV-1</td>
<td>0.2 μg</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0 μg</td>
<td>72</td>
<td></td>
</tr>
</tbody>
</table>

[0296] One important advantage of luciferase reporter is its short protein half-life. The enhanced expression could
also result from extended protein half-life and, if so, this gives an undesired disadvantage of the new gene. This possibility is ruled out by a cycloheximide chase ("CHX Chase") experiment (FIG. 14), which demonstrated that there was no increase of protein half-life resulted from the humanized Renilla luciferase gene.

To ensure that the increase in expression is not limited to one expression vector backbone, is promoter specific and/or cell specific, a synthetic Renilla gene (Rluc-
final) as well as native Renilla gene were cloned into different vector backbones and under different promoters (FIG. 13B). The synthetic gene always exhibited increased expression compared to its wild-type counterpart (Table 11).

<table>
<thead>
<tr>
<th>Vector</th>
<th>CHO</th>
<th>NIH-3T3</th>
<th>HeLa</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRL-tk, native</td>
<td>3,834.6</td>
<td>922.4</td>
<td>7,671.9</td>
</tr>
<tr>
<td>pRL-tk, synthetic</td>
<td>13,252.5</td>
<td>9,040.2</td>
<td>41,745.5</td>
</tr>
<tr>
<td>pRL-CMV, native</td>
<td>168,062.2</td>
<td>842,482.5</td>
<td>153,539.5</td>
</tr>
<tr>
<td>pRL-CMV, synthetic</td>
<td>2,168,129</td>
<td>8,440,306</td>
<td>2,532,576</td>
</tr>
<tr>
<td>pRL-SV40, native</td>
<td>224,224.4</td>
<td>346,787.6</td>
<td>85,325.6</td>
</tr>
<tr>
<td>pRL-SV40, synthetic</td>
<td>1,469,588</td>
<td>2,632,510</td>
<td>1,422,830</td>
</tr>
<tr>
<td>pRL-null, native</td>
<td>2,853.8</td>
<td>431.7</td>
<td>2,434</td>
</tr>
<tr>
<td>pRL-null, synthetic</td>
<td>9,151.17</td>
<td>2,439</td>
<td>28,317.1</td>
</tr>
<tr>
<td>pRL3b, native</td>
<td>12</td>
<td>21.8</td>
<td>17</td>
</tr>
<tr>
<td>pRL3b, synthetic</td>
<td>130.5</td>
<td>212.4</td>
<td>1,094.5</td>
</tr>
<tr>
<td>pRL3-tk, native</td>
<td>27.9</td>
<td>155.5</td>
<td>186.4</td>
</tr>
<tr>
<td>pRL3-tk, synthetic</td>
<td>6,778.2</td>
<td>8,782.5</td>
<td>9,685.9</td>
</tr>
<tr>
<td>pRL-tk no intron, native</td>
<td>31.8</td>
<td>165</td>
<td>93.4</td>
</tr>
<tr>
<td>pRL-tk no intron, synthetic</td>
<td>6,665.5</td>
<td>6,379</td>
<td>21,433.1</td>
</tr>
</tbody>
</table>

With reduced spurious expression the synthetic gene should exhibit less basal level transcription in a promoterless vector. The synthetic and native Renilla luciferase genes were cloned into the pGL3-basic vector to compare the basal level of transcription. Because the synthetic gene itself has increased expression efficiency, the activity from the promoterless vector cannot be compared directly to judge the difference in basal transcription, rather, this is taken into consideration by comparing the percentage of activity from the promoterless vector in reference to the control vector (expression from the basic vector divided by the expression in the fully functional expression vector with both promoter and enhancer elements). The data demonstrate that the synthetic Renilla luciferase has a lower level of basal transcription than the native gene (Table 12).

It is well known to those skilled in the art that an enhancer can substantially stimulate promoter activity. To test whether the synthetic gene has reduced risk of inappropriate transcriptional characteristics, the native and synthetic gene were introduced into a vector with an enhancer element (pGL3-enhancer vector). Because the synthetic gene has higher expression efficiency, the activity of both cannot be compared directly to compare the level of transcription in the presence of the enhancer, however, this is taken into account by using the percentage of activity from enhancer vector in reference to the control vector (expression in the presence of enhancer divided by the expression in the fully functional expression vector with both promoter and enhancer elements). Such results show that when native gene is present, the enhancer alone is able to stimulate transcription from 42-124% of the control, however, when the native gene is replaced by the synthetic gene in the same vector, the activity only constitutes 1-5% of the value when the same enhancer and a strong SV40 promoter are employed. This clearly demonstrates that synthetic gene has reduced risk of spurious expression (Table 12).

The synthetic Renilla gene (Rluc-final) was used in vitro systems to compare translation efficiency with the native gene. In a T7 quick coupled transcription/translation system (Promega Corp., Madison, Wis.), pRL-null native plasmid (having the native Renilla luciferase gene under the control of the T7 promoter) or the same amount of pRL-
null-synthetic plasmid (having the synthetic Renilla luciferase gene under the control of the T7 promoter) was added to the TNT reaction mixture and luciferase activity measured every 5 minutes up to 60 minutes. Dual luciferase assay kit (Promega Corp.) was used to measure Renilla luciferase activity. The data showed that improved expression was obtained from the synthetic gene (FIGS. 15A, B). To further evidence the increased translation efficiency of the synthetic gene, RNA was prepared by an in vitro transcription system, then purified. pRL-null (native or synthetic) vectors were linearized with BamHI I. The DNA was purified by multiple phenol-chloroform extraction followed by ethanol precipitation. An in vitro T7 transcription system was employed by prepare RNAs. The DNA template was removed by using RNase-free DNase, and RNA was purified by phenol-chloroform extraction followed by multiple isopropanol precipitations. The same amount of purified RNA, either for the synthetic gene or the native gene, was then added to a rabbit reticulocyte lysate (FIGS. 15C, D) or wheat germ lysate (FIGS. 15E, F). Again, the synthetic Renilla luciferase gene RNA produced more luciferase than the native one. These data suggest that the translation efficiency is improved by the synthetic sequence. To determine why the synthetic gene was highly expressed in wheat germ, plant codon usage was determined. The lowest usage codons in higher plants coincided with those in mammals.

Reporter gene assays are widely used to study transcriptional regulation events. This is often carried out in co-transfection experiments, in which, along with the primary reporter construct containing the testing promoter, a second control reporter under a constitutive promoter is transfected into cells as an internal control to normalize experimental variations including transcription efficiencies between the samples. Control reporter signal, potential pro-
motor cross talk between the control reporter and primary reporter, as well as potential regulation of the control reporter by experimental conditions, are important aspects to consider for selecting a reliable co-reporter vector.

[0303] As described above, vector constructs were made by cloning synthetic Renilla luciferase gene into different vector backbones under different promoters. All the constructs showed higher expression in the three mammalian cell lines tested (Table 11). Thus, with better expression efficiency, the synthetic Renilla luciferase gives out higher signal when transfected into mammalian cells.

[0304] Because a higher signal is obtained, less promoter activity is required to achieve the same reporter signal, this reduced risk of promoter interference. CHO cells were transfected with 50 ng pGL3-control (firefly luc+) plus one of 5 different amounts of native pRL-TK plasmid (50, 100, 500, 1000, or 2000 ng) or synthetic pRL-TK (5, 10, 50, 100, or 200 ng). To each transfection, pUC19 carrier DNA was added to a total of 3 μg DNA. Shown in FIG. 16 is the experiment demonstrating that 10 fold less pRL-TK DNA gives similar or more signal as the native gene, with reduced risk of inhibiting expression from the primary reporter pGL3-control.

[0305] Experimental treatment sometimes may activate cryptic sites within the gene and cause induction or suppression of the co-reporter expression, which would compromise its function as co-reporter for normalization of transfection efficiencies. One example is that TPA induces expression of co-reporter vectors harboring the wild-type gene when transfecting MCF-7 cells. 500 ng pRL-TK (native), 5 μg native and synthetic pRG-B, 2.5 μg native and synthetic pRG-TK were transfected per well of MCF-7 cells. 100 ng/well pGL3-control (firefly luc+) was co-transfected with all RL plasmids. Carrier DNA, pUC19, was used to bring the total DNA transfected to 5.1 μg/well. 15.3 μl TransFast Transfection Reagent (Promega Corp., Madison, Wis.) was added per well. Sixteen hours later, cells were trypsinized, pooled and split into six wells of a 6-well dish and allowed to attach to the well for 8 hours. Three wells were then treated with the 0.2 μM of the tumor promoter, TPA (12(14)-diacetoxyphorbol-13-myristate). Cells were harvested with 0.4 ml Passive Lysis Buffer 24 hours post TPA addition. The results showed that by using the synthetic gene, undesirable change of co-reporter expression by experimental stimuli can be avoided (Table 13). This demonstrates that using synthetic gene can reduce the risk of anomalous expression.

### Table 13: TPA Induction

<table>
<thead>
<tr>
<th>Vector</th>
<th>Rhu</th>
<th>Fold Induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRL-tk untreated (native)</td>
<td>184</td>
<td>4.4</td>
</tr>
<tr>
<td>pRL-tk TPA treated (native)</td>
<td>812</td>
<td>8.0</td>
</tr>
<tr>
<td>pRG-B untreated (native)</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>pRG-B TPA treated (native)</td>
<td>8</td>
<td>1.47</td>
</tr>
<tr>
<td>pRG-untrans (native)</td>
<td>132</td>
<td>4.36</td>
</tr>
<tr>
<td>pGL3-control (native)</td>
<td>195</td>
<td>4.36</td>
</tr>
<tr>
<td>pGL3-treated (native)</td>
<td>192</td>
<td>4.36</td>
</tr>
<tr>
<td>pGL3-control (final)</td>
<td>12,816</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### REFERENCES


[0344] All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification, this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details herein may be varied considerably without departing from the basic principles of the invention.
<210> SEQ ID NO 2
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of clone YG#61-6G01

<400> SEQUENCE: 2

tctaaacct

<210> SEQ ID NO 3
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 3

atgatgaaac gggagagaaa cgcatcatac ggcgcaagac ctcgtaaacc atctggaagc 60
ctgacgctc ctgagagctt atgtccgtct atctgtaasac atctcactt gctcaagcc 120
cctggttagt cttggtggca cgaacagtct ataatttata gatgtcagct 180
cttgtgccc agtcctcgtc taaatgggtct taacaattga acgatgtggt cagcatttctg 240
gctgagaca aaccccgctt tttaataccca gttgatgcg cttgtatact ccgctagatt 300
gtggccctgt cgaatgaact ttataatcga gcggaggtgct gcagagtcct ggtattagc 360
aaactctaa cctgctttaa ttaaagaaaac atctcgaaga aagttcggga atgtcgagct 420
cgtactaact tcgctaatgc catttcctt cttgatcgcct ctagaatctt ccggcgtgtc 480
agaacgctg caasctttat ttctgttatt aggcagccaga atccgtctaa cttcaagcct 540
cgctatattg atccagggag gcagagcgcg cgtatatttg cgttataggg cactagccgt 600
cgctcaag ggagtgacga ggtcacsacat aatctcgttg tcggcagctg tctatccttg 660
gaccacgctg tcggctaccc gttgatactc gcggaggtgc tcggcagctg tctatccttg 720
tttaacgtct cggatccttc tataactcttg ggtgtattcc gcggcgtgtc gcggcgtgtc 780
agtattctgc gttctagatga agaagctttat atggagggca tcggagacta cagagcgtgt 840
agcagcatc acgctctcct ctygcgttgg ttcctgagca atactccatt ccgctagtaa 900
tatgacgctg gcgtttctcgt ccagctgcgg tggcggcgtg ccgctattgc gaagaggtgg 960
gccgagctg ctcggcagct cgggtatattg cccgcttcgc gttgaggtgg tgcggctgct 1020
gagagccct ctgctacact ccctgatgct cggctactat ccaatctcgg ccgctagcgt 1080
cgctgaaccttt gggtaagg gcgttaagac gcggacgctg agacggcgaag actccgtggt 1140
cccaacctg cgggagac ggttgatatgc gcggcagctg cgcctggcag cctagcga 1200
aatgatctca ggcctagag acgctctgag gccggcagac gcggatccct 1260
agtatttgag cccctagaact cgtatcgtag gcggcagctg cgcctggcag cctagcga 1320
ttctacgac gtgctgtgg cggatcagag ctgctatcgc cggctattgc gagctacggtc 1380
atcgctagc tgccgctcgct ggtctagac gtttggggttg cggctactgt gcggcagctg 1440
tcttgctgtg aacaagcaag taaggaatt atcctagag aagctctaga cttgttgggc 1500
gacgctagc gtcttaccaaa gtttctgctcg ccggctgtgg cggctactgt gcggcagctg 1560
cggatcgtc gggcggagct cggctatcag ggtttccgctgg cgttggcag ctcggcagctg 1620

ggtggc

<210> SEQ ID NO: 4
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 4

atgatgaaac gggagagaaa cgcatcatac ggcgcaagac ctcgtaaacc atctggaagc 60
cctgacgctc ctgagagctt atgtccgtct atctgtaasac atctcactt gctcaagcc 120
cctggttagt cttggtggca cgaacagtct ataatttata gatgtcagct 180
cttgtgccc agtcctcgtc taaatgggtct taacaattga acgatgtggt cagcatttctg 240
-continued

gctgagaaaa acacacgctt ttctatctaa gttgattgccg ctttgaatatt cgccaatgatt 300
gtgcocctct ctaaatagaa ttaatatccga gacgaatccg gcaagcatcag ggtattacgc 360

aasctcacta ctgcttcttc tacccttsga atcttggacta aggtctctgagg aagctcgctt 420
cgtaactact ttcacacgct cattatattc tttgatatcc gtagaataat ccagacgctgt 480
gagacggtgc caaactttat ttctcgttat agogacggtaaaaatcctagcata ctctaaagct 540
cctatcttgg atccagaacgg gcgaagctgg ccgtatatttt agcttacagc cccacccgtg 600
cctgtcaag gcctgatagcc gacctaaaggct aatcctgtg tccgctttgat tgtcgcctgct 660
gaccaacgggg tggctgctgta gttgactctc ggtctgctgta ctctcctgctt ccctgccctc 720
tttcaacgtct cttgctttttt ctcattcgct ggttttcttc cttctgctttc ggtgctgatc 780
atgttccgtct gtctgctagcct cagacgctt gcaagcgccac ttcacagacact cgaagctgct 840
agctgatcact gtttctttttt gctcctgttg gctgtgcattt cttctgctttc ggtgctgatc 900
tatgacgactct ccagctctgcct cagacgctgct cttctgctttc cctccctttgc taaaaaggtg 960

gcgaagctgct gctgcaagccgt cctatatttt ccgctttgat tgtcgcctgct 1020
gacgacactct ctgctactat ttcctttggct cttctgctttc ccctccctttgc taaaaaggtg 1080
ggtgactcct cgatgtggct cttctttttt ctcctttttt cttctgctttc ggtgctgatc 1140
ccaaataaag cctgctctttt gttcatagtc gttcttcttt cttcttcttt cttctgctttc 1200
aatgttggag ccacagccact accatcgtact cttctgctttc ggtgctgatc 1260

ggcttactcc ttcagctgaaa ctttctttggct cttctgctttc ggtgctgatc 1320
tataaagct ttcgctgctt ccctctctct cttcttcttt cttctgctttc ggtgctgatc 1380
attctgcat cttctggctgcc gttatctcttc gttcttcttt ttcctctttt cttctgctttc 1440

<210> SEQ ID NO: 5
<211> LENGTH: 1626
<212> TYPE: DNA
<220> ORGANISM: Artificial Sequence
<222> FEATURE: OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 5

atgatgaaacc gcgaagaaaaa gctgatctac ggcgggctag cactgcgtcc ctgtggagac 60
cctcagctgg gttgctgctc ctcatagactcgt cctactctcc atagagcagct gctccagctg 120
cctgtggagc cttctgctct cggagacgct cttgataaag catctctctcc agctgtcgtg 180
cgtggcctcc aaacggtcgct taaagctgct ccacagcact cagagctgtgct gctcgcttcg 240
ggtggagata aagctgcttt ttttcatctcc gttgctgctc ctgcagcatct gcggtaggtg 300
gtgcocctct ctaaatagaa ttaatatccga gacgaatccg gcaagcatcag ggtattacgc 360

aasctcacta ctgcttcttc tacccttsga atcttggacta aggtctctgagg aagctcgctt 420
cgtaactact ttcacacgct cattatattc tttgatatcc gtagaataat ccagacgctgt 480
gagacggtgc caaactttat ttctcgttat agogacggtaaaaatcctagcata ctctaaagct 540
ctgcatttgg aatccagcga gcaggtggcc gcacatcttg gcttcttcttg ccacactctgt 600
tggctaaag gttgcagaca gctacaaag aatactgtag tgtgcttcttg ccacacccca 660
gaacctccttg tgggtactca attgatccct ggctgacact tggctgtgta tttgccttcc 720
tttcaagcct tgggtttttc tatccacctgt ggcattttca tggctgcctg tgtgctgact 780
agttttcgcg cggccacca agagccttac ctggaagctca tactagaact cagagttgct 840
tctgctga caagcctccat tgcctataag ttccggtagca ttcctccttt ggttgccaa 900
tagctctctga gcagctcctgc tgaacctgct gcgtggcggct ctcttttcgc caagaagaagt 960
gcggcgctcg ctgctcaagc gctggacact cctggtctcc cgggtgtctg tgtttgact 1020
gagccaccc ttcgcacacat ccctagctgg cgtgacgacgt ttaaactcttg tagcctgctg 1080
cggctgcgacc ctctgactggt tccgaaatgc gcgacacgct gacgcygcaaa agcctctggc 1140
cacaaactgg cgctgtgaaa ggtgcttgag ctctcaagag ctctctagtg ctaagctgac 1200
aatgctgagc cccatccacag agctattgtg ggtgcttgggt ggttgcaatag cggcgctactc 1260
ggtttattag atgagacgcc acoccttcat gttgtcgctg gctataaagct atgtattagg 1320	taacaaagct ctcaagcctc gcgcagctaa gctggaaagatt ttttgcctgaa gacacctttg 1380
attcgccagcg gcggctgcgt ggtgatctca gactgggaag atggcaggtg tcctagagcc 1440	ctttggtgct aaccagctgcc caaagaggt aatcataagct attctaaagct ttcattgcgg 1500
gagcctgtct cccactatag atatcgtcgt gcggctgcctg gcttcgctgaa ttcctccct 1560
cgcacactct gccgaagat cactgcgtaa gcggtgcgta acacattgct gcaaaagact 1620
ggcgcgc 1626

<210> SEQ ID NO 6
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: OTHER INFORMATION
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 6

atgatgaaac gcggaaaaag atggcgatgt ac gcgggcaac cacagc gacgtaagag 60
cagctcgctgt ctcagctgca ctccgcttca acacgcct gccctcaggt 120
cgcggcaagc cacacccctc gccgaaaaag attccttcctc gcagctctgt 180
gacggacgcct gcgggtcgt tgcgtcgtg cagccttcct gcgtggcctg cggcggtgtt 240
ggtgacagat cccgcacccct ttctttccct gcgtgctagt ggtctttgcct gcctttggt 300
gctggagctcg gcgggtcgt tgcgtcgtg cagccttcct gcgtggcctg cggcggtgtt 360
gagctctcttc ttcagcataa attccttcct gcagctctgt 420
gagctctcttc ttcagcataa attccttcct gcagctctgt 480
cgcgggcaac cacagc gacgtaagag 540
gacggacgcct gcgggtcgt tgcgtcgtg cagccttcct gcgtggcctg cggcggtgtt 600
gacggacgcct gcgggtcgt tgcgtcgtg cagccttcct gcgtggcctg cggcggtgtt 660
gagctctcttc ttcagcataa attccttcct gcagctctgt 720
gagctctcttc ttcagcataa attccttcct gcagctctgt 780
gacggacgcct gcgggtcgt tgcgtcgtg cagccttcct gcgtggcctg cggcggtgtt 840
-continued

tcagtctaca atgtcctca cgtcatttgc ttcctgagca aatctccttt ggtgtcgaag 900
tatgatgta gcaagctctgc tgaaggttgc tgtgggctgt ctctttgagcc caaagaagt 960
gccagagttc ctcgtaagct ctcgaacccc ctcgctattc gctgcgtgtt cttggttact 1020
gacaggacct ctcgcaact ccatagcctg ccagagcagtc taaaggtccg tcagctgtt 1080
gcgcgactgc ctcctatggc tgcgaagact gcgcagcagtc ggcgcaagca aagccacggc 1140
ccaatcaagc tgcgattcatt ggtttataaag ggccttacac tgtcataaag ctcgtcgaac 1200
atgcagaggg ccacatasaga agcattgtat gtgagttgct ggtttcatag cgcagcactc 1260
aggtctatac atggaaaagc acactttcat gttgcagcgt tcatcaagaaga atgattaag 1320
tacaagctc tcacagtcct cccagcagca ctcgaagaaaa gttgtgtagaa gaaacccctt 1380
atcgcagacgc tgcgctgatt tgtatataa gactgggaag cttggtgtat gtcagacgcc 1440
cttattctga aacacacttg gcagagctcg acgtgatacg acgtatacg cttttgggc 1500
gcgcgctgag tcgcacacac atatctgctgc ggtgggcgttc gcttttggcg ttctcactca 1560
cgcgcgagc cgcgatctac cactctgtaaa gaaattgctga aagccaccttc cgcacagag 1620
ggcgcgcgcgcgcg 1626

<210> SEQ ID NO 7
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 7

atgatgaaac gcgcaggaac cgytgacact gcgccgaacc ccttgtcata acgtggaaac 60
cgctcaagct gttgagattg atctctgaga ctcgctattc cgtctactgc gccctgagca 120
cgctctgagg ccgggagaga cggacgctcc ctcgctacaa acttttttga agtcactgtg 180
cgttgtgacgc aaagcttaca taattttggt gcactataac gccttattgt ggagttttgt 240
gcgtgcgacgc aactctcatt ctattatctgc gtaagctgct cttgtgacat ccgctattgc 300
gtcgcgagcc tggagatcag tctcatacgc ggtactgctg gttacttgc gggtaattacg 360
aaaacctaaac agttccattc acatcagaacct ctaatagaga aagcagttgga agtaaccgtc 420
cgctatcaagt ccctctctgc ccagaggtttcc atgctatatt atgctacmcgc ccagctgtt 480
gagagctcct ctaactaatc ttgagttgcg acgtattacgc taatgcgttga tttcagcggcc 540
tgcatttga tcatcagcgc gccacagcgc gctatattgg tgcacgctgc gcaacctgctg 600
ttcgctagag ggtaggtgcc gactcaccag aatcagtttg gttacttgcg caggtctcgc 660
gacgctgtgc tgggtacaca attgactattt ggcgtgatctg tggatacagt ttcgctcttc 720
ttcagctgtc tgggttccct tattacagtt gcagatttc gaagcggctg ggtgctacot 780
tgtttgacct gctgccagca aagcgctctgc tgaatcgcct caaagatcct gcagttcgct 840
tcgggtgtgc agctccctgc agtcattttg ttgccagca acatctcctt ggtgttcgaag 900
tatgacgacgc gcagctcgctc tcaggtgtggg tggtgcactt gcacctttgc gcaggtgcgg 960
gcagagttgcc ctgctagcct ccgtgacctc ggtgtgtcct tggagttctt tcgygtgact 1020
gagagactt ctcgataact ccataggttt cggagacgct taaaggtcgg tagcctggttg 1080
cgcggtactgc ctcttatggc gcgcaagagc gcggcagcagtc ggcgcaagca aagccacggc 1140
-continued

ttcagtagag gctattttta cctgtagc ACCATGCAAG gggctgtcgt

<210> SEQ ID NO 8
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 8

atgcatgcg gtagcttcag ggcggaggg cagatgagct gctgacttc acagatgagct 60
ctcagctccg tctaggactgt ccagctactcc acagatgagct ccttactactcc 120
cctgactacc cagtggtggt gctggtgccg gctggtgccg cctggtgccg cctggtgccg 180
cctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg 240
gctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg 300
gctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg gctggtgccg 360
acagatgagct cccgacagatgagct cccgacagatgagct cccgacagatgagct cccgacagatgagct 420
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 480
ccagatgagct cccgacagatgagct cccgacagatgagct cccgacagatgagct cccgacagatgagct 540
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 600
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 660
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 720
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 780
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 840
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 900
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 960
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1020
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1080
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1140
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1200
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1260
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1320
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1380
cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg cctggtgccg 1440
-continued

```
ttggtggtga accaccggg caagagatct actgactaggg aggtctgctg tcaatgagcg
1500
gagcgcgctg ctcaccacca aatctgcgg gcggcgcgct gcgtcgcgta aaccatcaca
1560
cggacgtaa cggtagatg caactgtaaa aatctgcgg ccagaaaaaagt
ggcgc
1620

<210> SEQ ID NO 9
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQ ID NO: 9

atgatgaaac gcaaaaaga cgtgaactac gcggccagaac cactgcatac actgaaagac
60
cctacgctg gtgaagctct ttcgagagca ctcgatgca aatgtaaact cactoaagca
120
cctgggtggc cggatgcct cctcagaaaa tatttttgc aagttttgtg
180
cggttggcgc aagctctaca tattttttgc tcagaaaaag agcggttttgt gcagattttgt
240
ggcgaacta aacactgcgt ctttttttt ctagtctgctt tgtaaagtc tgcagttgtt
gtcggtcctg tgaattgaaac ttcagttatgt gcagatgtgg gcagataac
300
ttggatcata cagatgaggc gcaagctggcc gctatttttt gtctcctgag caccacgtttg
360
ctgctcaagag tggtagctca gactccacag aatctctgtg tgcgttggag cgacgctctc
gacgctgct ggggtcctgt gcggctgctg caagctctgc atgcggctgc
tttaacgcctt gcggtgccctg ttctcctgactg cgggtggttt gcagttttgt
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
420
gacgctgctg cgggtcctgt gcggctgctg caagctctgc atgcggctgc
tttaacgcctt gcggtgccctg ttctcctgactg cgggtggttt gcagttttgt
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
480
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
540
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
600
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
660
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
720
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
780
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
840
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
900
ttgcagcttg gcggcttttt atgcgtcctg cggagctggcc gcgtcctttg ttcggttttt
960
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1020
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1080
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1140
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1200
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1260
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1320
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1380
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1440
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1500
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1560
gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc gcggagctgc
1620
ggcgc
1626
```
<210> SEQ ID NO 10
atgatgaagc gtgagaaaaa tgtgttatat ggtctgtgac cattgcatac tctgagggat 60
ttgctctgct gagaaatgct gtttccgccc ttggtccaga agaagccatt ccagagcagtct 120
ttgctctgac gtgcttgcag tgtctctcct agctacaagg aatcttttgag ggccaaacgg 180
ttgctctgct caaacctgca caactctgcc tattagtgta aagagctcgtgt ccttactcgc 240
gccgaaaaca atactcggttt ctttatctcct gttctgtgcct cttggtatat tgttgattgc 300
gtgctctgac tcaacgagac ctaacctcctg gttgaaatcg tgaacaagct ggcacacctc 360
aacccagaga tgtcttctac cactaaaat atcttgagaag aggctgtagg ggtocaaagc 420
cgcccaaat ttataaacgc tacattatgc ttggaactctg tggaaacact ctaacaggttc 480
agatgtatgc ctaactttctc cagccgcttc tgtgctggca accctgcaaa tttaaacaac 540
ttgcaacctgc aacgtgcttc acaagctgctg gcaactcttg gtactccctt taccagcttcg 600
ttgccaaag ggctgacatc aaccacatcg acacatgtgg tgtgtctctgt gcaagctctc 660
gaccccgtcct acggactccta acagctctcc aagctgaagct gccctgagct tgtgtctctt 720
ttcacacttg tgtgggttcc ttcactcctg ggttaactaa ctggtggcgt cggctgctt 780
agttgcctgc gttctgacac gggcccttct tggaaagcata tocgagttctg ctggtgcct 840
tctctctata atctgtcagac ctoctaccgct cttttgctca agagccctct ggggcaacaa 900
tacgattgtg ctaacgctgct tgtggtctgtg tcgggtgcgc ctoctaccgc caagagatgc 960
gcggaggtgg cgcctaaacc atgtgaccctg cctggtacctgc tgggtgttctt gcggtgcaac 1020
gaactctca ggcctacatc caacactcttg cggagacgct taaagagccct tcttttggg 1080
cggtcaccac cactgtgctgc tcggccacatt gctatgctcg agactgtgtaa gcagctggcc 1140
cctacgacgg tgtggctgcct gtctagcacc ggcctggagc gctacaggg tgtgtgtaat 1200
aagctgacag ctccagcag aaccctgtagc gcacctggcct ggctgcacc gccgctgggg 1260
gggtactcag aagagatac ggaattttcat gttggtgcac gttaataagg aactgtaaacc 1320
tatcagggtga gcaagctggc ctoctgcgaga tggagggcct tcttgggttt cagagcgagt 1380
atcggagat ggtccgcatct cggccagagg ccgttcgaat cggccgctgtc 1440
tcgcttgct acagccgact caggagatct gtcgctggag aatgtgtgtaa ttctttgggg 1500
gacgctgctc gcttaccacac atacttggct gttgctgcac gtctttacctca 1560
cgttaacgtga ctcggtggtc gacccgaaac caaagtctgta aggcaactgtt ggagaagcc 1620
ggccgt 1626

tggcgccttctttctcgcctgtgctttgctcccctttggtgagctgtctttgctggagttcaagcaacagc
<table>
<thead>
<tr>
<th></th>
<th>Amino Acid</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ttgactgccc ggcgaatgct gtttcgcgc ttggcgcagc acagcactct gcacaaagtct</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>ttgatgagcgc tggatggcc tgaactctct gactaaagaac aggctgcttg gcagaacgcctg</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>ttgctgttgc acacgtgctg cccactgttg gctgactgctg gctactgtgc ttgtatgccttgc</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>gcagaaacaa atactgctctt cttatctct gcacatcttg cctgtatagtt ggtgtatctcc</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>gttgcccagc tcaacagagc ctacatctct gcagactgtg gtaagggctg ggcaagctctt</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>aagcaacaag ctcggcttac ccaatatat accttgacca actgtctgaa ggtctgaaagc</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>cgcggccaaat tgcattaatcg tgcattaatc tgcgattcttg gggaagacat tactyyggtgc</td>
<td>480</td>
<td></td>
</tr>
<tr>
<td>gaacttggcag ctatattcat ccagcagttacc tggatgctgca actggccaaat tttaaaaaa</td>
<td>540</td>
<td></td>
</tr>
<tr>
<td>ttcggcctctct gcagctgctg gcagctgctg gctagctctg gtagcctctc tactacgctg</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>ttgcgcaag gttgcagcag aacccctttcg acctttgtctgt tggcctcttt gcagctcttc</td>
<td>660</td>
<td></td>
</tr>
<tr>
<td>gatcatacatt ggcgaatgctg gtaaggtctg gtagcctcag gtagcctactg gtaaggtcttg</td>
<td>720</td>
<td></td>
</tr>
<tr>
<td>tccccatgtt cccagctttgg cgcttttcttt ctttt tgtcctgcc gccgtgcattt</td>
<td>780</td>
<td></td>
</tr>
<tr>
<td>atgctctgctg gttgggcttcc tgggaaggttt ttgaagacttc gcgtatggaag</td>
<td>840</td>
<td></td>
</tr>
<tr>
<td>ttgatgctata ctatgctcttg gcagcttctg ttcgattcttg gcagctttggc ggcagctttggc</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>tgcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>960</td>
<td></td>
</tr>
<tr>
<td>gctgagggcgc gcgcgttactgc gctgattcacc ctggggcttt gcgggttctgg gcgggttctgg</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>ggaactactgc ggcagcttctg cccagctttgg cccagctttgg cccagctttgg cccagctttgg</td>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>cgcgcctttg cccagctttgg cccagctttgg cccagctttgg cccagctttgg cccagctttgg</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>ctaagcaagc gttgggcttcc tgggaaggttt gcagtctctgc gtagcctactg gtagcctactg</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>gctgatgctata ctatgctcttg gcagcttctg ttcgattcttg gcagctttggc ggcagctttggc</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>ggcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1320</td>
<td></td>
</tr>
<tr>
<td>ggcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>atgcagagtcc gcgcgttactgc gctgattcacc ctggggcttt gcgggttctgg gcgggttctgg</td>
<td>1440</td>
<td></td>
</tr>
<tr>
<td>ggcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>ggcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>gcgggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>ggcggttatttcttctctg tgcggttatttcttctctg tgcggttatttcttctctg</td>
<td>1680</td>
<td></td>
</tr>
</tbody>
</table>

**SEQUENCE:** 12

- **FEATURE:**
  - **OTHER INFORMATION:** Sequence of a synthetic luciferase
-continued

cagcacaaca tgtgttctac gactagatct atttgcgaac aagtctgtag cgtccaaagc 420
cgcaacact ttaataaagc ttaatcactc tggacacgct tcggagatat taaggtttgc

gacactctgc ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 540
	tggctatccg acctctggaga aacatggtgct gctatccgtt tcgacagcgg tactacgtgc

cgctaaagg ggcctagcga gaccccataa aacattgtgct cgggtctgtcat ccctgctccctc 660
ggcctactac ggtgactcctc ggtgacccgct tttgcctctag cgtgctctcct 720
ttcgactcctt gtcgctccac gattaacttgg ggtcatcatt gggctgcatt  780
	agttgctggc gttctgctgta ggaggctttt tcggacgacca ttaaagagta tgaagctgcg 840

tgcgctctct aatgctctcg cgtgctcgct ttggcgtactt tcgacccgct ccttaagcttc 900

tcagcactgt ctctccgctgg tgaagctgct cgggctgctgg ccccaacgctg cagaggctgc 960

gtcgagcttg ccgctcaaac cttgaacttg ccagcatttc gtcgggttct ccgctctctc 1020

gaatcactca ggcgtcatatt ctcaactcttc ccggctgagct ttaaagcgcc cccttttgccgc 1080

gctgctacct ccatcctgag cggcctaatg gaaagtctg aaccgctgac gaaacctggcc 1140
cctacaacaag tgggaggcct gttgatcctaa cgggctgctg ggaagctggtt 1200

cgctgtaagc ttcgctccct cggcctgctg cggcctgctg cggcctgctg 1260

ggccactctg gcggcactct ggcgaactgtt ccgctgttctt tgggctgatt 1320
tacacaagga gcggagcagc gatcctgtag cggggagctgg cttcctggaa aaacctcttg 1380

tcagcactgt ctctccgctgg tgaagctgct cgggctgctgg ccccaacgctg cagaggctgc 1440

ttcgactcctt gtcgctccac gattaacttgg ggtcatcatt gggctgcatt  1500

gaacccggaac gcccctgctg ggggctgctg cttgctgtca cgcctactaa cggcctgctg 1560
gctgctcctt cgtgactctg cctccctgcgc gccttgctgc cggcctgctg cggcctgctg 1560

gacactctgc ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 1620

ggtgcctgc

gcagactctg ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 1626

<210> SEQ ID NO 13
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 13

atgatgaagc ggagaatgaa cgtctctatg ggcctgctgc cttgctcctg tttgagat 60
ttgactgcgg gcaaatctgc tgtgtctgtat tcgctctagc gtccttcatttc gcgccctgagc 120
ttgctgctgg tgtgagcctg tgaatctcttg agtccacacc gtttttttgt ggcgaacggctg 180
tggtctgctgg cgtctgaactc ttaattctgtgc tcaatactgaa aagacgtctgtct cttccattgtg 240
gcagcacaaca ataccgcttt cctccactga gctatcgccg catgttatct cggatgcatc 300
gtgcctgctgc tcacgctgag ctcattctgg cgcagctctg tgaatcctgc gggctttctct 360
gacactctgc ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 420
cgcaacact ttaataaagc ttaatcactc tggacacgct tcggagatat taaggtttgc

gttcgactcctt gtcgctccac gattaacttgg ggtcatcatt gggctgcatt  480

gaatcactctg ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 540
gctgctacct ccatcctgag cggcctaatg gaaagtctg aaccgctgac gaaacctggcc 600
cggaagcctg gacggctcct ggtgacccgct tttgcctctag cgtgctctcct 660

cgctaaagg ggcctagcga gaccccataa aacattgtgct cgggtctgtcat ccctgctccctc 720
ttcgactcctt gtcgctccac gattaacttgg ggtcatcatt gggctgcatt  780
	agttgctggc gttctgctgta ggaggctttt tcggacgacca ttaaagagta tgaagctgcg 840

tgcgctctct aatgctctcg cgtgctcgct ttggcgtactt tcgacccgct ccttaagcttc 900

tcagcactgt ctctccgctgg tgaagctgct cgggctgctgg ccccaacgctg cagaggctgc 960

gtcgagcttg ccgctcaaac cttgaacttg ccagcatttc gtcgggttct ccgctctctc 1020

gaatcactca ggcgtcatatt ctcaactcttc ccggctgagct ttaaagcgcc cccttttgccgc 1080

gctgctacct ccatcctgag cggcctaatg gaaagtctg aaccgctgac gaaacctggcc 1140
cctacaacaag tgggaggcct gttgatcctaa cgggctgctg ggaagctggtt 1200

cgctgtaagc ttcgctccct cggcctgctg cggcctgctg cggcctgctg 1260

ggccactctg gcggcactct ggcgaactgtt ccgctgttctt tgggctgatt 1320
tacacaagga gcggagcagc gatcctgtag cggggagctgg cttcctggaa aaacctcttg 1380

tcagcactgt ctctccgctgg tgaagctgct cgggctgctgg ccccaacgctg cagaggctgc 1440

ttcgactcctt gtcgctccac gattaacttgg ggtcatcatt gggctgcatt  1500

gaacccggaac gcccctgctg ggggctgctg cttgctgtca cgcctactaa cggcctgctg 1560
gctgctcctt cgtgactctg cctccctgcgc gccttgctgc cgcctactaa cggcctgctg 1560

gctgctacct ccatcctgag cggcctaatg gaaagtctg aaccgctgac gaaacctggcc 1620

ggtgcctgc

gcagactctg ctaaatcact tgggcctgtat tcgacgagca aacctcacaac cttaaaacct 1626
-continued

gatccagct aagggcttca gctgattctt ggtgtaeccc ccctttgctca ctggccttccc tgggtggc 720

ttcocattg cgggtactca tattacattg ggtcatttta gttgggtcct ctggcattg 780

agtttcccgt ctgggtcacta gggggtcttc ttctggactca tcaaaagcta gtagctcctg 840

aatggatttac gctgctgcag ctgctctgtt ttttctca gtagccactc agctgtgacc 900

tcgagattc ctctcaaggt ctagttggtg tgggtgctgc ctgctgcc gcaggggtcc 960

gtctgattg ccggccttac cttggaactg cttggtgctt gctggcttga ccggcctgcc 1020

gagctatcat tccgattcct gcggatcagt ctagaaggg ctttctggg 1080

cgtcctactc cactagtcca tcgtacatgc gcgatacctg cgaattgttaa ggttcttggc 1140

cctaaactgg ttgggtggag ctgggtacta ggcgtatttg gggccaggg ttattgctat 1200

aacgtgcaag cctcctggga ggcgtatcag gacgcaacgt cttgggtactc tggctatgtt 1260

ggatattac gcgaagagct gcgtatttac gcgtattcag cttcactacag gcgatacaaa 1320

tcaggggtga ggcccattgt gcggctgcag ttgattgtga ttctgcttgg aaatatccgc 1380

attctagtgg aggtcgattg gcgtgctgct ggtgattgag ggggctagct ggtctctgtg 1440

ttcggtcag ccgatcactg cccggttgct cccgcagtt gcgttatgtt tttccagcttg 1500

gacgcttgga gcgtacttta ctggctgatt ggcgtattgg gttttgctga tccgattcct 1560

cggccactgc ctggcccttac actgctcagc gcgggttgtt gaacatggtg ggagaagggc 1620

gggtgt 1626

<210> SEQ ID NO 14
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 14

agatgagag gttgaaaaa ttgcaactat ggccgtgac ccctcactcc ttgaggagat 60

ttgatctgct ggcagaatct tttggtctt ccctcagcag actocacttt gctcacaacc 120

ttggctatgt tcgtgggcta ttcacattag agctcacaag gtgtttttga gcaagcgctg 180

ttgctggtcct gccctcacta caagtgtgct gcgtcactga acggggtgtg gtagattttg 240

gtctggactc cttcacttca gcgctgcag cattgatatag cgttatgtat 300

gttgctgctt gcgtattgac cttactcttc gcgtcactga gtagctgatt gcgtttggct 360

agccacacag tttcggatt ctcggattgt cgtcttactg gcgtctcaag cgtctgctcag 420

cgctgacttc tttataattt gagaacatcgt cagagattg ggtcattatt ttttatttattt 480

gattttcg ctttcttttact ctttctttact ctttctttact ctttcttttact ctttcttttact 540

cctcacttg accttctctct cgggtacttt actggcaactgc gtagatttgc gcgtctcagt 600

cctcttcgcgg atgcctggtg accctctttg ggcttctttt cgtcttactg 660

gatccttctt ggttcttttag cttgcttact gcgtctcagt gcgtcttactgcgccttctcct 720

ttcatttct catttctttt catttctttt catttctttt catttctttt catttctttt 780

agtttggcct cttggtcacta gggggtcttc ttctggactca tcaaaagcta gtagctcctg 840

agttcctcag ctgctgcag ctgctctgtt ttttctca gtagccactc agctgtgacc 900

tcgagattc ctctcaaggt ctagttggtg tgggtgctgc ctgctgcc gcaggggtcc 960
continuation

gctgaagtg ccggccaaac gccttgaatt ccagggcttc gttggtgctt cgggctcaac 1020
gaatocca gcggctatt tcgcttcttc cctgatggcg ttaagacggc ctctttgggc 1080
cgctgcgcta ccctcagcgc tgctagagtc gctgatcggc aaacgttgtaa ggcgtttgggc 1140
cctaccaag ttggcgctgt cgctgctaaca gcggcaagcg tttaataaggt ctaattaat 1200
aacgtgcaag cctaccaagcc ggccatgac gcggacggtc gttggtcatt tgtgtgattt 1260
ggatattcag agaaacagta gctattttaac gctggtggtcc gttccacgga gctgataacaa 1320
taccagggta gccaggtgcc tcgagctgag ttgagcaggt tttcttggaac aatacctgac 1380
attccgctag tcgctgctgt cgctccatct cccttgagcg acgctggaac gctctctgtgct 1440
tttgctgttc aacgccgctg taaagacatt cccgccacgc aacgtgttatg ttactggtc 1500
gactcggtga ccgctctacaa tttctggctgt gcgctgcgct gcgttggtgca ctcacttccct 1560
cgtaacgtaa cccggtaaaat ttcctgctgaag gcgcgctgtta aacaatttgt ggagagccgc 1620
ggcctg 1680

<210> SEQ ID NO 15
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 15

atgatsggaag gcggagaaaaa tgtcatctat ggccctgagc ctctccaccc tttggaggat 60
ttgacggtgcc ggacatcttc gttcttgctc ttcgctgacgg actctatttt gcggcagcc 120
tgggtcggcg tgtggcgggga tgaatctgtt cagctacacgg agttggtttaga ggctaagcgc 180
ttgctgcggtt ctctccaccc gctagctgctc tcagctgtgtggtgactcc 240
gctgaacaa ataccccgtct tttccactgct gctatgcgct catggtatat cggctgaactc 300
gtggctggcg tcaaacgagc ctaaactctcc gacgccactgc tgaataagct gcggctgctc 360
agcagccaga tttctggcgt cccagagcatt ctctgcaacct cagctcggct ggcctagaacgc 420
cgctacacg tttgctcttc ctctctactg ttgagccagc atctacccacct cctccaccoc 480
gaatcttggc atgttctttc ctctctactg ttgagccagc atctacccacct cctccaccoc 540
ctctccactg cccctggccag acaaggtgctt gctagcgagc tccactgctg 600
cctcgaagaag gcggcagaaaaa gttcaatgttc tggctctggtc aaccagcagc 660
gatccagctc gcctcgtgct ccctcgtgct ccctcgtgct ccctcgtgct 720
ttcctgctt gctgtgcttt tttacacttt gttactttta tttgctgtct gcggcagc 780
atgctctgg gcgggtggttt gccagcctgc tttgctgtct gcggcagc 840
agctgatcac gttcctgctgc tttgctgtct gcggcagc 900
tacactgctt cctcctgctgc tttgctgtct gcggcagc 960
gctgatcag gcggcagaaaaa tggctgctgc tttgctgtct gcggcagc 1020
aatcctttcc gcctctactt gcctctactt gcctctactt gcctctactt 1080
cgggctctgc gcctctactt gcctctactt gcctctactt gcctctactt 1140
cgcagccaga tttgctctgc tttgctgtct gcggcagc 1200
aacgtgatac gcggcagaaaaa gcggcagaaaaa gcggcagaaaaa gcggcagaaaaa 1260
-continued

ggatattcgc acgaagatga gcctttttac gtcgctggtc gttacaagga gcgtgatcgg 1320
tcaagaagtc gcgtcgtggt tcagctgtgag ttggagaggag ttctttgag tatccatcgg 1380
atcctcgagc tcctcgctcag ccgacctctct gatctgtgag ccgacctgctc gccttcgct 1440
tctgctggtgc aagcgctcggt caattaaatt acgcacccag aagctgtatgt tttaacgctg 1500
gacggttgc gcctactaa acgtggctggt gcctgtggttc gttttttgtg gtcgcagcccc ctctacgct 1560
cgtaacgctaa cagcagttaa taacgcgcaag gacgtggttga acaatgttgtt ggtggagggccc 1620
gcgcggc 1626

---

<210> SEQ ID NO 16
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 16

tagatgaagc tgagaataaa tttcaatcct ggtccctgac ctctccattct tttggaggat 60
tcgcaagcgc ggcagacagt cgctgtctgt ctcgcaagcgc ccctccatttt gcctgaagcgc 120
tgtgtgatga tgggtggagc tgaatcttttt tagctacaag ggtttttttat ggcacacgctc 180
tggtcgtgtc gtcacgtaag ccatttcgac tcaaatagta acagcagtcag tgtgttgctgt 240
gcctgaaacca atacgcgatt cctctacattagac tcagccggtg cctgttagat ccgctatcgc 300
gttcgtgagc tcctcgagag ttcacatcctcc gaaattctggct gtatctcgat ggttatactct 360
agaccaaacag ttgctctgct cactcgaaat tattttgagc aagctctgga gttcccaagc 420
cgacccaact tttaataacg ttcacacgct tggccacagtg tgggaaatgta tcacgctgct 480
gtagctttggc ctcacactct ctcgctgtct ctcgcaagcgc ccctccatttt gcctgaagcgc 540
cctccagcc acgttcgagg acaagttgca gaccaacattc gtaaccaggg tacroctagag 600
cctccagcc gactttactc gcacccgaatt aacatctcag aggcttgctg ccatgctctc 660
gacacgctg cagctacgct ggtgttctgct ggttccacgct tttttttcag ccagttcttggc 720
tcaggtcctg ccgctttctcc tatactctgtt ggtatctcctg tttgctcttc ctcagcgggt 780
agttggcgg cttggctcttc gaggctctta ttcacaagcta tcagagattc tagatcgggg 840
gagttgacttc gctggatcag cgtatctgag tttaatgcac agggccagct cgttgtcagcg 900
tacagcttt gttctacagct gtaatctggt tgggttcgag cttcagctggc ttaaggggtcg 960
gctgaagttc cgcccaagcc cttggattct cctgggtcatt ggttggcatt cgtggccacc 1020
gaatctttc ggctttactt cttatctcct gcgtagattgt tttaaggggct ctttttgaccg 1080
gcgctgacttc cactctagcag tgcacagatg ggtatcagc caacatcgttta ggttggctccg 1140
cggcttacagt gggcagttcatt gttctaacaag gacccatagag tcagcagacat ctgcatcagacatg 1200
acagcatagct crccacaggt gccagcagtag gctgctggtgc agggttctttt cttgctgaggc 1260
ccatagacag cagcagatga gcatttttcc gctgctggtgc gttctaacaag gacccatagag 1320
taacatagct crccacaggt gccagcagtag gctgctggtgc agggttctttt cttgctgaggc 1380
atcgccagtc gcgctcttct gctctggagg cctggagacact gctctctcgtct 1440
ctcctctcag ccgctcttct gctctggagg cctggagacact gctctctcgtct 1500
agaagctgat gcacattcag ttcagcctgag ggcctgctcag ttctctctgc 1560
cytaacgtc aagcacaat taccgcgaag ggcctgttga aacaattttg ggagaaggccc 1620
ggcgg 1626

<210> SEQ ID NO: 17
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<222> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 17
atgctgaagc gtgacaaaaa ttgcatctat ggcctggagc ctctcactcc ttggtgagat 60
ctgccctgcc gcgaactgct gtttctgtgc ttccgacaac gctctcatcttc ggcctacaag 120
taggctggag tggagcgtga caaagagtga gtcgctcctg tagctaatgc 180
tgcgctggtc gtttcatctcc ctctacaccca ggcctggagc tgggtctctgtag 240
ttcgtgata gcctttctct ctttcttct cccttcagctc ttcctttcttct 300
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 360
ctcttcctgg ttcctttcttct ccctttcttct ccctttcttct ccctttcttct 420
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 480
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 540
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 600
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 660
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 720
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 780
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 840
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 900
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 960
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1020
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1080
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1140
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1200
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1260
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1320
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1380
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1440
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1500
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1560
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1620
ttcctttcttct ccctttcttct ccctttcttct ccctttcttct ccctttcttct 1686
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 18

atgataaagc gtagagaaaaa tgttcatctat ggccccgggc ctcctccactc cttggaggtt 60
ttgactcgc ggaattgtgt cggtgtgtggt ctcctccagc actctctctt ggccagcggc 120
ttggtcgagc tgcttcggga tgaatcttttg agttttttgga ggcgaagcgtc 180
ttgctggcgc aggccctccg cacttgggca tcaagtgctga agccgctggt cgttacgtct 240
gctgaaaaca atacccggttt ctcctccactc gtcgactgcg cattgatatct cgtaaggtac 300
gtgcgtcggc tcaagccagga ctcctccccc ggcgaactctg ttcgaactctt ggttacgtct 360
aagccacagc ttgcttcctac cactaatgaa aattcggagga agtttcagcc 420
cyccacagc ttaattacgt attaatattc ccctgagcact ctgcgtgcgtgac 480
gaatctttgc ctaatattct cttctgtatat tcaagcggca caactcaaaaa cttttaacc 540
cgcacattcgc acacggttgcga acaagctttga gcgttacgcgtg gcgacgctggcg 600
cgcacagc ggtttcggag caacactgca aacattcttg gctggtctgt gaattctc 660
gacatcgaac agcagcactc gttggtctcttt ctttcctggtc gctggtctcttc 720
ttccacggtt gttgtctcttc tatctattgtt gttgtctcttt cgtggtttttc 780
atggttggct cttctctcccc ggtgtctcttt cgtggtttttc 840
agtttcgttgc gggtttggtt gggcagctt acaagcagaaggc 900
tacaattctgc ctcctccgag gccttctgctt tggggtctcttc 960
gcgttacggt ccggccacagtttacggtttcagtggctggctgggc 1020
ggcgttttat ccagactttc ggggatggtt tttaagacgg ctttctggggc 1080
cggggttcttc ctcctctcttc gcgaagggcgt gcgttctcttc gcgttctcttc 1140
cgcagagcagttggagttg gtttactgtc aaccccttctc ggcttctcttc gcgttctcttc 1200
aaggtctaa cggagctttcg cggctcttctg tggggtctcttc gcgttctcttc gcgttctcttc 1260
ccgggtcatc cggactttctg tggggtctcttc gcgttctcttc gcgttctcttc gcgttctcttc 1320
tacaattctgc ctcctccggag gccttcttctg tggggtctcttc gcgttctcttc gcgttctcttc 1380
attcggttct cccagtctgct cgggtggtgc cggggtctcttc gcgttctcttc gcgttctcttc 1440
ttctctctgc cccagtctgct cgggtggtgc cggggtctcttc gcgttctcttc gcgttctcttc 1500
gacagctttcg ccggctttctg tggggtctcttc gcgttctcttc gcgttctcttc gcgttctcttc 1560
cggggtttatt ctcctccgag gccttcttctg tggggtctcttc gcgttctcttc gcgttctcttc 1620
ggcggt 1626

<210> SEQ ID NO 19
<211> LENGTH: 933
<212> TYPE: DNA
<213> ORGANISM: Renilla reniformis

<400> SEQUENCE: 19

atgacttggc aagtttttga tcggaaaaac ccgggagggg tggttacttg gcctgactgtg 60
ttggagccgt tataactgtg gatttttctt gttctacattc ttaatattc tggggtcgc 120
aaactatcag aaatctgctg tttttttttc ctgtgaaccc cgggctttc ctggatattg 180
cgcctacttg tggcactatt cgggtgtcttt tataactggtc tttttttttc ctggggtcgc 240
-continued

```
ATG GCA AAT CAG GAG CTC TGC TCA GAG GAG CAC AAG CCT

CTTA GCT GTG TCG TCA GGC GGT TCA GAG GAG AAG CTC

TTA AAT GAC CCA CCA TCC TCA GGT GGT TCA CTC

GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG

CTT GCA GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC

GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG

```

<210> SEQ ID NO 20
<211> LENGTH: 933
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 20

```
ATG GCT TTA GCA AGT GCT CCA GCG GCG CTC GCT GCG

TGG GCT GGT TGG CAG GGA GCT GCT TAA TTA CAA CAA

AAG CAC GAG GCG GCT GCT CAG GGG GCG GCT GCT CTA

GAG CTC TGA CAG GCG CAG GCG TGC TGC TGC TGC TGC

ATG GCG AGT CTT GCT GCT GCT CTC TCA GCT GCT GCT

CTG CAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG

TGG GCT TTA GCA AGT GCT CCA GCG GCG CTC GCT GCG

GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG

CTT GCA GAC GAC GAC GAC GAC GAC GAC GAC GAC GAC

GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG

```

SEQ ID 0 21

LENGTH 933

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 21
atggtcca aagggatca cgccacgccca cggacaacgc tgcagcctgg gcccgctgg 60
tgactgtct gcaaacatt gaaacgctct ggcactcctca ctggaactac agcactcag 120
tgaacagccc ggaaggcttc agattttcttg cagtgaaacag cggcctagc attgtgtgcct 180
tagagaagc ctgctactct cagaggcttg gtcctctctg tccctgtga ctcagcctta 240
tagttgaagt cggctagcag cggggatggc tcattgccct tcctggtcata gtagctiggct 300
tctacagctt gcgtgcagtc gtgaacctt ccgagaaaaa ttattaattg ggcccacgac 360
tgggggcctt ccctgccttc tcaactactcc ttcagacacc aagcaacagct caagcctccc 420
tgcctgatg ctagtggtcg gaagtggtac ggcctctggg aggctgctg ccctgatcag 480
gaggtattct cctgaggcag gcgcagaagag ggagagaaaa ttgagctttgc gaataactctc 540
tccgctgaaca ccctgcttcc aagcaacagct gggaggagac cggacagccttc cgagctggct 600
tccattctgg aagcctccaa ggaaggagcgc gggatttggc gcgtacctct cctgagtgcct 660
cggagagcc cccgggtcatt ggagagcagc cggacagcttc cagatcgtg cggccccttc 720
acggctaccc ttgggcctgg cggccgcttc ccaagcactg tccagagcct cggagatctgc 780	
tctcttctca aagcatctttg cggcagagct aagaggtcct ctaaagcagct gttgcttgaag 840
gttgagggcc gccctctcag cccgagagac ggtccccagtag aatgggttat gtaactcaag 900
gagcagctgg aagcaggtct gcgaacagcag cag 933

<210> SEQ ID NO 22
<211> LENGTH: 933
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 22
atggtcca aagggatca cgccacgccca cggacaacgc tgcagcctgg gcccgctgg 60
tgactgtct gcaaacatt gaaacgctct ggcactcctca ctggaactac agcactcag 120
tgaacagccc ggaaggcttc agattttcttg cagtgaaacag cggcctagc attgtgtgcct 180
tagagaagc ctgctactct cagaggcttg gtcctctctg tccctgtga ctcagcctta 240
tagttgaagt cggctagcag cggggatggc tcattgccct tcctggtcata gtagctiggct 300
tctacagctt gcgtgcagtc gtgaacctt ccgagaaaaa ttattaattg ggcccacgac 360
tgggggcctt ccctgccttc tcaactactcc ttcagacacc aagcaacagct caagcctccc 420
tgcctgatg ctagtggtcg gaagtggtac ggcctctggg aggctgctg ccctgatcag 480
gaggtattct cctgaggcag gcgcagaagag ggagagaaaa ttgagctttgc gaataactctc 540
tccgctgaaca ccctgcttcc aagcaacagct gggaggagac cggacagccttc cgagctggct 600
tccattctgg aagcctccaa ggaaggagcgc gggatttggc gcgtacctct cctgagtgcct 660
cggagagcc cccgggtcatt ggagagcagc cggacagcttc cagatcgtg cggccccttc 720
acggctaccc ttgggcctgg cggccgcttc ccaagcactg tccagagcct cggagatctgc 780	
tctcttctca aagcatctttg cggcagagct aagaggtcct ctaaagcagct gttgcttgaag 840
gttgagggcc gccctctcag cccgagagac ggtccccagtag aatgggttat gtaactcaag 900
gagcagctgg aagcaggtct gcgaacagcag cag 933
<210> SEQ ID NO: 23
<211> LENGTH: 543
<212> TYPE: PRT
<213> ORGANISM: Pyrophorus plagiophthalmus

<400> SEQUENCE: 23

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His 1   5   10   15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg 20  25  30
Lys His Ser His Leu Pro Gln Ala Leu Val Asp Val Phe Gly Asp Glu 35  40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Cys Leu Leu Ala Gln 50  55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys 65  70  75  80
Ala Glu Asn Asn Lys Arg Phe Phe Ile Pro Ile Ile Ala Ala Ala Trp Tyr 85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu 100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Glu Ile Val Phe Cys Thr 115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe 130 135 140
Ile Lys Arg Ile Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys 145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala 165 170 175
Asn Phe Lys Pro Leu His Tyr Asp Pro Val Glu Gin Val Ala Ala Ile 180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gin Thr 195 200 205
His Gin Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Ala 210 215 220
Gly Thr Gin Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe 225 230 235 240
Phe His Ala Phe Gly Phe Ser Ile Asn Leu Gly Tyr Phe Met Val Gly 245 250 255
Leu Arg Val Ile Met Leu Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys 260 265 270
Ala Ile Gin Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ala Ile 275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser 290 295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val 305 310 315 320
Ala Glu Val Ala Val Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly 325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Gly Asp 340 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala 355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val 370 375 380
Gly Glu Leu Cys Val Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn 385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gln Trp Leu His 405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu His Phe Tyr Val Tyr 420 425 430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro 435 440 445
Ala Glu Leu Glu Ile Leu Lys Asn Pro Cys Ile Arg Asp Val 450 455 460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala 465 470 475 480
Phe Val Val Lys Glu Pro Gly Glu Ile Thr Ala Lys Glu Val Tyr 485 490 495 500
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Tyr Leu Arg Gln Gly 505 510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr 515 520 525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ser Ser Lys Leu 530 535 540

<210> SEQ ID NO 24
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Sequence of clone YG#61-6G01

<400> SEQUENCE: 24
Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1   5   10   15
Pro Leu Glu Asp Leu Thr Ala Gly Met Leu Phe Arg Ala Leu Arg
20  25  30   35
Lys His Ser His Leu Pro Gin Ala Ala Val Asp Val Val Gly Asp Glu
35  40  45   50
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
50  55  60   65
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
65  70  75  80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gin Ile Val Phe Thr Thr
115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe
130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ser Ile Trp Asp
145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165 170 175
---continued---

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

<210> SEQ ID NO 25
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 25

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1      5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
20     25  30
Lys His Ser His Leu Pro Glu Ala Ala Val Val Gly Asp Glu
35     40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
50     55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Val Val Ser Ile Cys
65     70  75  80
Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
85     90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100    105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr
115    120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gln Ser Arg Thr Asn Phe
130    135 140
Ile Lys Arg Ile Ile Leu Leu Asp Thr Val Glu Asn Ile His Gly Cys
145    150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165    170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gln Val Ala Ala Ile
180    185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
195    200 205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Val
210    215 220
Gly Thr Gln Leu Ile Pro Gly Val Val Leu Val Tyr Leu Pro Phe
225    230 235 240
Phe His Ala Phe Gly Phe Ser Ile Thr Leu Gly Tyr Phe Met Val Gly
245    250 255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gln Glu Ala Phe Leu Lys
260    265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275    280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
290    295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305    310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325    330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Arg Asp
340    345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
355    360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
370    375 380
-continued

| Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn  |
|---|---|---|---|---|---|
385 | 390 | 395 | 400 |   |
| Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His  |
405 |   |   |   |   |
| Ser Gly Asp Phe Gly Tyr Asp Glu Asp Glu His Phe Tyr Val Val  |
420 | 425 |   |   |   |
| Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gln Asp Glu Val Ala Pro |
435 |   | 440 |   | 445 |
| Ala Glu Leu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val  |
450 |   | 455 |   | 460 |
| Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala |
465 |   | 470 | 475 |   | 480 |
| Phe Val Val Lys Gln Pro Gly Lys Glu Ile Thr Ala Lys Gly Val Tyr  |
485 |   | 490 |   | 495 |
| Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly  |
500 |   | 505 |   | 510 |
| Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr |
515 |   | 520 |   | 525 |
| Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Asl Gly Gly  |
530 |   | 535 |   | 540 |

<210> SEQ ID NO 26
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 26

| Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His  |
|---|---|---|---|---|---|---|---|---|
1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg  |
20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| Lys His Ser His Leu Pro Gin Ala Leu Val Asp Val Val Gly Asp Glu  |
35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 |
| Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln  |
50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 |
| Ser Leu His Ann Cys Gly Tyr Lys Met Ann Asp Val Val Ser Ile Cys  |
65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 |
| Ala Glu Ann Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr  |
85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 |
| Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu  |
100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 |
| Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gin Ile Val Phe Thr Thr  |
115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 |
| Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe  |
130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 |
| Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys  |
145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 |
| Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala  |
165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 |
| Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile  |
180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 |
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr 195 200 205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Val 210 215 220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe 225 230 235 240
Phe His Ala Phe Gly Phe Ser Ile Thr Leu Gly Tyr Phe Met Val Gly 245 250 255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gln Glu Ala Phe Leu Lys 260 265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val 275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser 290 295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val 305 310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly 325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Arg Asp 340 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala 355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Glu Val 370 375 380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Val 385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Asp Gly Trp Leu His 405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val 420 425 430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Gly Ser Glu Val Ala Pro 435 440 445
Ala Glu Leu Glu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val 450 455 460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Leu Pro Ser Ala 465 470 475 480
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr 485 490 495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Val 500 505 510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr 515 520 525
Arg Lys Glu Leu Leu Lys Glu Leu Leu Glu Gly Lys Ala Gly 530 535 540

<210> SEQ ID NO: 27
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 27
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Met</td>
<td>Lys</td>
<td>Arg</td>
<td>Glu</td>
<td>Lys</td>
<td>Asn</td>
<td>Val</td>
<td>Ile</td>
<td>Tyr</td>
<td>Gly</td>
<td>Pro</td>
<td>Glu</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Ser</td>
<td>Tyr</td>
<td>Lys</td>
<td>Glu</td>
<td>Phe</td>
<td>Ser</td>
<td>Tyr</td>
<td>Lys</td>
<td>Met</td>
<td>Asn</td>
<td>Asp</td>
<td>Val</td>
<td>Ser</td>
</tr>
<tr>
<td>85</td>
<td>90</td>
<td>95</td>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Met</td>
<td>Ile</td>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Ala</td>
<td>Glu</td>
<td>Ser</td>
<td>Tyr</td>
<td>Ile</td>
<td>Pro</td>
<td>Asp</td>
</tr>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asn</td>
<td>Ile</td>
<td>Leu</td>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Ala</td>
<td>Asn</td>
<td>Val</td>
<td>Glu</td>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Pro</td>
<td>Asn</td>
<td>Phe</td>
<td>Ile</td>
<td>Ser</td>
<td>Arg</td>
<td>Tyr</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Asn</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Ser</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Thr</td>
<td>Gly</td>
<td>Leu</td>
<td>Pro</td>
<td>Lys</td>
<td>Glu</td>
<td>Val</td>
<td>Met</td>
<td>Gln</td>
</tr>
<tr>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asn</td>
<td>Ile</td>
<td>Cys</td>
<td>Val</td>
<td>Arg</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>pro</td>
<td>Arg</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Thr</td>
<td>Gly</td>
<td>Val</td>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Ala</td>
<td>Val</td>
<td>Tyr</td>
<td>Leu</td>
<td>Pro</td>
<td>Phe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Leu</td>
<td>His</td>
<td>Ala</td>
<td>Phe</td>
<td>Gly</td>
<td>Ser</td>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
<td>Tyr</td>
<td>Phe</td>
<td>Met</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
<td>Met</td>
<td>Phe</td>
<td>Arg</td>
<td>Phe</td>
<td>Asp</td>
<td>Glu</td>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>270</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ile</td>
<td>GLN</td>
<td>Asp</td>
<td>Tyr</td>
<td>GLu</td>
<td>Val</td>
<td>Arg</td>
<td>Ser</td>
<td>Val</td>
<td>Ile</td>
<td>Asn</td>
<td>Val</td>
<td>Pro</td>
<td>Ser</td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Ser</td>
<td>Lys</td>
<td>Ser</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Leu</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Arg</td>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
<td>Ala</td>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>GLu</td>
<td>Ala</td>
<td>Ala</td>
<td>Lys</td>
<td>Arg</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
<td>Pro</td>
<td>Gly</td>
<td>Ile</td>
<td>Arg</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>GLy</td>
<td>Leu</td>
<td>Thr</td>
<td>GLu</td>
<td>Ser</td>
<td>Thr</td>
<td>Ser</td>
<td>Ala</td>
<td>Asn</td>
<td>Ala</td>
<td>Leu</td>
<td>Ser</td>
<td>Leu</td>
<td>Arg</td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLu</td>
<td>GLy</td>
<td>Ser</td>
<td>GLy</td>
<td>Ser</td>
<td>Leu</td>
<td>GLy</td>
<td>Arg</td>
<td>Val</td>
<td>Thr</td>
<td>Pro</td>
<td>Leu</td>
<td>Met</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ile</td>
<td>Ala</td>
<td>Asp</td>
<td>GLu</td>
<td>THR</td>
<td>GLy</td>
<td>Ala</td>
<td>Leu</td>
<td>GLy</td>
<td>Pro</td>
<td>Asn</td>
<td>GLu</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLy</td>
<td>GLu</td>
<td>GLu</td>
<td>Cys</td>
<td>Ile</td>
<td>GLy</td>
<td>GLy</td>
<td>Pro</td>
<td>Val</td>
<td>Ser</td>
<td>Lys</td>
<td>GLy</td>
<td>Tyr</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>395</td>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Val</td>
<td>GLu</td>
<td>Ala</td>
<td>Thr</td>
<td>Lys</td>
<td>GLu</td>
<td>Ala</td>
<td>Ile</td>
<td>Asp</td>
<td>ASP</td>
<td>GLy</td>
<td>TCR</td>
<td>Leu</td>
<td>His</td>
</tr>
</tbody>
</table>
\begin{verbatim}
Sar Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val
405  410  415
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Gly Ser Gln Val Ala Pro
420  425  430
Ala Glu Leu Glu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
435  440  445
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
450  455  460
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
465  470  475  480
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
485  490  495
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
500  505  510
Arg Lys Glu Leu Leu Lys Gln Leu Leu Glu Lys Ala Gly Gly
515  520  525
530  535  540
\end{verbatim}
210  215  220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
225  230  235  240
Phe His Ala Phe Gly Phe Ser Ile Thr Leu Gly Tyr Phe Met Val Gly
245  250  255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys
260  265  270
Ala Ile Gin Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275  280  285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
290  295  300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305  310  315  320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325  330  335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Arg Asp
340  345  350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
355  360  365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gin Val
370  375  380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385  390  395  400
Aan Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
405  410  415
Ser Gly Asp Phe Gly Tyr Tyr Glu Asp Glu His Phe Tyr Val Val
420  425  430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gin Val Ala Pro
435  440  445
Ala Glu Leu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
450  455  460
Ala Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
465  470  475  480
Phe Val Val Lys Gin Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485  490  495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500  505  510
Val Arg Phe Val Asp Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515  520  525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly
530  535  540

<210> SEQ ID NO 29
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: Luciferase
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 29
Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
Lys His Ser His Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu  
35                                           40  
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln  
50                                           55  
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys  
65                                           70  
Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr  
85                                           90  
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu  
100                                          105  
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr  
115                                          120  
Lys Aan Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Aan Phe  
130                                          135  
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys  
145                                          150  
155                                          160  
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala  
165                                          170  
175                                          180  
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile  
185                                          190  
Leu Cys Ser Ser Gly Thr Thr Leu Pro Lys Gly Val Met Gin Thr  
195                                          200  
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Val  
210                                          215  
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe  
225                                          230  
235                                          240  
Phe His Ala Phe Gly Phe Ser Ile Thr Leu Gly Tyr Phe Met Val Gly  
245                                          250  
255                                          260  
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys  
265                                          270  
275                                          280  
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val  
285                                          290  
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser  
295                                          300  
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val  
305                                          310  
315                                          320  
Ala Glu Val Ala Ala Leu Asn Leu Pro Gly Ile Arg Cys Gly  
325                                          330  
335                                          340  
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Arg Asp  
345                                          350  
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala  
355                                          360  
365                                          370  
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gin Val  
375                                          380  
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn  
385                                          390  
395                                          400  
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Asp Gly Trp Leu His  
405                                          410  
415                                          420  
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val  
425                                          430  
435
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro
435  440  445
Ala Glu Leu Glu Glu Ile Leu Leu Asn Pro Cys Ile Arg Asp Val
450  455  460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
465  470  475  480
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485  490  495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500  505  510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515  520  525
Arg Lys Glu Leu Leu Lys Glu Leu Leu Glu Lys Ala Gly Gly
530  535  540

<210> SEQ ID NO 30
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 30
Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
  1   5   10   15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
 20   25   30
Lys His Ser His Leu Pro Glu Ala Leu Val Asp Val Val Gly Asp Glu
 35   40   45
Asn Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
 50   55   60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
 65   70   75   80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
 85   90   95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100  105  110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Glu Ile Val Phe Thr Thr
115  120  125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe
130  135  140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys
145  150  155  160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165  170  175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile
180  185  190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gin Thr
195  200  205
His Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Val
210  215  220
Gly Thr Gin Leu Leu Ser Gly Val Thr Val Leu Tyr Leu Pro Phe
225  230  235  240
<SEQ ID NO 31>
<LENGTH> 542
<TYPE> PRT
<ORGANISM> Artificial Sequence
<FEATURE> OTHER INFORMATION: Sequence of a synthetic luciferase

<SEQUENCE> 31

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5           10   15
Pro Leu Glu Asp Leu Thr Ala Gly Met Leu Phe Arg Ala Leu Arg
20  25        30
Lys His Ser His Leu Pro Gin Ala Leu Val Asp Val Val Gly Asp Glu
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
    50  55  60
Ser Leu His Asn Cys Gly Tyr Met Asn Asp Val Val Ser Ile Cys
    65  70  75  80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
    85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
    100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr
    115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gln Ser Arg Thr Asn Phe
    130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys
    145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
    165 170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Gln Val Ala Ala Ile
    180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
    195 200 205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Val
    210 215 220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
    225 230 235 240
Phe His Ala Phe Gly Phe Ser Ile Thr Leu Gly Tyr Phe Met Val Gly
    245 250 255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gln Glu Ala Phe Leu Lys
    260 265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
    275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Tyr Leu Asp Ser
    290 295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
    305 310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
    325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Asn Ile His Ser Leu Arg Asp
    340 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
    355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
    370 375 380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
    385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
    405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val
    420 425 430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Glu Val Ala Pro
    435 440 445
<table>
<thead>
<tr>
<th>1</th>
<th>Met</th>
<th>Met</th>
<th>Lys</th>
<th>Arg</th>
<th>Glu</th>
<th>Lys</th>
<th>Asn</th>
<th>Val</th>
<th>Ile</th>
<th>Tyr</th>
<th>Gly</th>
<th>Pro</th>
<th>Glu</th>
<th>Pro</th>
<th>Leu</th>
<th>His</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Pro</td>
<td>Leu</td>
<td>Glu</td>
<td>Asp</td>
<td>Leu</td>
<td>Thr</td>
<td>Ala</td>
<td>Gly</td>
<td>Glu</td>
<td>Met</td>
<td>Leu</td>
<td>Phe</td>
<td>Arg</td>
<td>Ala</td>
<td>Leu</td>
<td>Arg</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Lys</td>
<td>His</td>
<td>Ser</td>
<td>His</td>
<td>Leu</td>
<td>Pro</td>
<td>Gln</td>
<td>Ala</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Val</td>
<td>Gly</td>
<td>Asp</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Ser</td>
<td>Leu</td>
<td>Ser</td>
<td>Tyr</td>
<td>Lys</td>
<td>Glu</td>
<td>Phe</td>
<td>Phe</td>
<td>Glu</td>
<td>Ala</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
<td>Ala</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>Ser</td>
<td>Leu</td>
<td>His</td>
<td>Asn</td>
<td>Cys</td>
<td>Gly</td>
<td>Tyr</td>
<td>Lys</td>
<td>Met</td>
<td>Asp</td>
<td>Val</td>
<td>Val</td>
<td>Ser</td>
<td>Ile</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Ala</td>
<td>Glu</td>
<td>Asn</td>
<td>Thr</td>
<td>Arg</td>
<td>Phe</td>
<td>Phe</td>
<td>Ile</td>
<td>Pro</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Trp</td>
<td>Tyr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Ile</td>
<td>Gly</td>
<td>Met</td>
<td>Ile</td>
<td>Val</td>
<td>Ala</td>
<td>Pro</td>
<td>Val</td>
<td>Asn</td>
<td>Glu</td>
<td>Ser</td>
<td>Tyr</td>
<td>Ile</td>
<td>Pro</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Leu</td>
<td>Cys</td>
<td>Lys</td>
<td>Val</td>
<td>Met</td>
<td>Gly</td>
<td>Ile</td>
<td>Ser</td>
<td>Lys</td>
<td>Pro</td>
<td>Gln</td>
<td>Ile</td>
<td>Val</td>
<td>Phe</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Lys</td>
<td>Asn</td>
<td>Ile</td>
<td>Leu</td>
<td>Asn</td>
<td>Lys</td>
<td>Val</td>
<td>Glu</td>
<td>Val</td>
<td>Glu</td>
<td>Ser</td>
<td>Arg</td>
<td>Thr</td>
<td>Asn</td>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td>135</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>Ile</td>
<td>Lys</td>
<td>Arg</td>
<td>Ile</td>
<td>Ile</td>
<td>Leu</td>
<td>Asp</td>
<td>Thr</td>
<td>Val</td>
<td>Glu</td>
<td>Asn</td>
<td>Ile</td>
<td>His</td>
<td>Gly</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>Glu</td>
<td>Ser</td>
<td>Leu</td>
<td>Pro</td>
<td>Asn</td>
<td>Phe</td>
<td>Ile</td>
<td>Ser</td>
<td>Arg</td>
<td>Tyr</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Asn</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>Asn</td>
<td>Phe</td>
<td>Lys</td>
<td>Pro</td>
<td>Leu</td>
<td>His</td>
<td>Phe</td>
<td>Asp</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Gln</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Ile</td>
</tr>
<tr>
<td>185</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>Leu</td>
<td>Cys</td>
<td>Ser</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Thr</td>
<td>Gly</td>
<td>Leu</td>
<td>Pro</td>
<td>Lys</td>
<td>Gly</td>
<td>Val</td>
<td>Met</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>His</td>
<td>Gln</td>
<td>Asn</td>
<td>Ile</td>
<td>Cys</td>
<td>Val</td>
<td>Arg</td>
<td>Leu</td>
<td>Ile</td>
<td>His</td>
<td>Ala</td>
<td>Leu</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Tyr</td>
</tr>
<tr>
<td>215</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>Gly</td>
<td>Thr</td>
<td>Gln</td>
<td>Leu</td>
<td>Ile</td>
<td>Pro</td>
<td>Gly</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
<td>Tyr</td>
<td>Leu</td>
<td>Pro</td>
<td>Phe</td>
</tr>
<tr>
<td>230</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>235</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>Phe</td>
<td>His</td>
<td>Ala</td>
<td>Phe</td>
<td>Gly</td>
<td>Phe</td>
<td>His</td>
<td>Thr</td>
<td>Leu</td>
<td>Gyl</td>
<td>Tyr</td>
<td>Phe</td>
<td>Met</td>
<td>Val</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
-continued

Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys
  260   265   270
Ala Ile Gin Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
  275   280   285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Tyr Asp Leu Ser
  290   295   300
Ser Leu Arg Glu Leu Cys Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
  305   310   315   320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
  325   330   335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gin Ser Leu Arg Asp
  340   345   350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
  355   360   365
Lys Ile Ala Asp Arg Glu Thr Gly Ala Leu Gly Pro Asn Gin Val
  370   375   380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
  385   390   395   400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gin Trp Leu His
  405   410   415
Ser Gly Asp Phe Gly Tyr Asp Glu Asp Gln Ile Phe Tyr Val Asp
  420   425   430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gin Val Ala Pro
  435   440   445
Ala Glu Leu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
  450   455   460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
  465   470   475   480
Phe Val Val Lys Gin Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
  485   490   495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
  500   505   510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
  515   520   525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly Gly
  530   535   540

<210> SEQ ID NO: 33
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 33

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Glu Gly Pro Glu Pro Leu His
  1    5   10   15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
  20   25   30
Lys His Ser His Leu Pro Gin Ala Leu Val Asp Val Val Gly Asp Glu
  35   40   45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gin
  50   55   60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
65

Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Ala Trp Tyr
85

Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100

Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr
115

Lys Asn Ile Leu Asn Lys Val Leu Glu Val Glu Ser Arg Thr Asn Phe
130

Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys
145

Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
160

Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gln Val Ala Ala Ile
175

Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
190

His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr
205

Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
220

Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly
235

Leu Arg Val Ile Met Phe Arg Arg Phe Asp Glu Ala Phe Leu Lys
250

Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
265

Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Tyr Asp Leu Ser
280

Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305

Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
320

Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Ser Leu Arg Asp
335

Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
350

Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
365

Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
380

Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
405

Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val
420

Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro
435

Ala Glu Leu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
450

Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
---continued---

**SEQ ID NO:** 34  
**LENGTH:** 542  
**TYPE:** PRT  
**ORGANISM:** Artificial Sequence  
**FEATURES:**  
**OTHER INFORMATION:** Sequence of a synthetic luciferase

<400> SEQUENCE: 34

```plaintext
465  Phe Val Val Lys Gln Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr  
470  495  490  495
475  Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly  
480  500  505  510
485  Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr  
490  515  520  525
495  Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly Gly  
500  530  535  540
```

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His  
1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg  
20  25  30
Lys His Ser His Leu Pro Gin Ala Leu Val Asp Val Val Gly Asp Glu  
35  40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln  
50  55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys  
65  70  75  80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr  
85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Gly Ser Tyr Ile Pro Asp Glu  
100  105  110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gin Ile Val Phe Thr Thr  
115  120  125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe  
130  135  140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys  
145  150  155  160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala  
165  170  175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile  
180  185  190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gin Thr  
195  200  205
His Gin Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr  
210  215  220
Gly Thr Gin Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe  
225  230  235  240
Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly  
245  250  255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys  
260  265  270
Ala Ile Gin Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
Ser Leu Arg Glu Leu Cys Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Ser Leu Arg Asp
Glu Phe Lys Ser Gly Ser Gly Ser Gly Val Thr Pro Leu Met Ala Ala
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
Ser Gly Asp Phe Gly Tyr Asp Glu Asp Gly His Phe Tyr Val Val
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro
Ala Glu Leu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Leu Pro Ser Ala
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
Arg Lys Glu Leu Leu Lys Gln Leu Leu Glu Lys Ala Gly Gly

<210> SEQ ID NO 35
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 35
aagccagccc aagcttaggc ctgagtgcc

<210> SEQ ID NO 36
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 36
cattaatcctcc cacatcctccc tgtgacaat taatcagcc ctcq
-continued

<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 37

tataatgtga ggaatgaga gggataaca atttacaca 40

<210> SEQ ID NO 38
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 38

atggaatgtt accgtacca atatgaaata tttgtaaaat 40

<210> SEQ ID NO 39
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 39

aaatgcctaa tgaaatttca aaaaaaaaaa aaggaattc 40

<210> SEQ ID NO 40
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 40

gatatacagc ttatacaga cgctgacacc gggattata 40

<210> SEQ ID NO 41
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 41

tagaaaaagc ctcgcqcgqc cgctgttca gtcqgtt 37

<210> SEQ ID NO 42
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 42

aactgactgaa atagscg 17

<210> SEQ ID NO 43
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 43
gcgcgcggg ccttttttctt tatactctc gcgcgcgcg

40

<210> SEQ ID NO: 44
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 44
gtatcgataa gctgtgtatc gaaatcttt ttttttttttt

40

<210> SEQ ID NO: 45
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 45
gaattatat cgaatcttt ttttttttt ttttgaatct

40

<210> SEQ ID NO: 46
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 46
ttttgaatct ttcagcatg atttatcact atttttctat

40

<210> SEQ ID NO: 47
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 47
tgtcgtatgt aacattcacc cactagcttt ttttttata

40

<210> SEQ ID NO: 48
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 48
tgcsatttcc tccacatctc cggagcgatg attaatgatc

40

<210> SEQ ID NO: 49
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 49
tgcsatttcc tccacatctc cggagcgatg attaatgatc
<210> SEQ ID NO: 50
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 50

ggasacagga tccatgag aasgcgsaa agacgtgtg

<210> SEQ ID NO: 51
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 51

cctaccgcc caccctgag aaccagccag cccactgacc

<210> SEQ ID NO: 52
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 52

gcctggtgag tgctcttcag gcacagtggc aacacagtc

<210> SEQ ID NO: 53
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 53

acccacac gcacagtgcg gcacgtgtgg gacacagcag

<210> SEQ ID NO: 54
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 54

ccctctcaaa aaagaattt tgaagaacac tgtcgtgtg

<210> SEQ ID NO: 55
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 55

coccaagcc tccataaatg tgggtaaaa atgaaagctg

<210> SEQ ID NO: 56
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 56

tggtgagcat tggtgagag aataacactc gtttctttat
40

<210> SEQ ID NO: 57
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 57

tctgttaatc gttggtgttat aacttgaggt gcgtggtgcc
40

<210> SEQ ID NO: 58
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 58

cctgtgaatg aactttactc ccacagttag gttgtaagg
40

<210> SEQ ID NO: 59
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 59

ttatgggtat tagcaactc ccaatgctct ttaactccaa
40

<210> SEQ ID NO: 60
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 60

aasactcttc aataaaggtct tgggaactca gttcctgact
40

<210> SEQ ID NO: 61
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 61

aacttccac aaccgcctcat tatcctggt tactctgcgaaaa
40

<210> SEQ ID NO: 62
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 62

acctccacgg ctgtgagacg ctcctactctctctctcg 40

<210> SEQ ID NO: 63
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 63
ttcagcgt ggtaaatcg ctaaatccga gcccttgcat 40

<210> SEQ ID NO: 64
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 64
tttgatacg tcgagcaagc gcogcgatt ttgtgctct 40

<210> SEQ ID NO: 65
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 65
cggcaccgc tggtttgct aaaaagttgca tgcagactca 40

<210> SEQ ID NO: 66
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 66
cccagaatc tggttgctgt tggtaaacgc ttcctgaccct 40

<210> SEQ ID NO: 67
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 67
cgtgtggtga ctaaatgtgt cccgtgctgt cgtgtgcgtg 40

<210> SEQ ID NO: 68
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 68
<210> SEQ ID NO 69
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 69
tgtctgccc tttcttttac gcttttgtt tcctctttac 40

<210> SEQ ID NO 70
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 70
cctggtctat ttcattgtcg gctttgctgt ctctctgttt 40

<210> SEQ ID NO 71
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 71
cgctgcttgcc accaacaagc ctctttgaa gcttttcaag 40

<210> SEQ ID NO 72
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 72
actaagagtt gcttcttcgt acaacaactc ctctagcctcat 40

<210> SEQ ID NO 73
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 73
tttgtctctg acacaatcct ctttctttga caaagtatgt ctg 37

<210> SEQ ID NO 74
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 74
agcagcttgc gtcagcttgt ctttgctgtc atcctcctt 37

<210> SEQ ID NO 75
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 75
tggccaaaga agtggccagag tgcagctgta apgctcgtga 40
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 75

ccccctgt gtcggtgctg gtttggggt ggtgagaq  40

<210> SEQ ID NO 76
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 76

acacctgca acacacatag cttgagagac ggtttaagt  40

<210> SEQ ID NO 77
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 77

cctctgagct gggctcctgag actctctta tggctgcaaa  40

<210> SEQ ID NO 78
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 78

gacgcgac gcggagacgc gazagaacgt gggcacaat  40

<210> SEQ ID NO 79
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 79

cacagcctgt tgaatttat taacgccctt atggtctca  40

<210> SEQ ID NO 80
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 80

agagtagct gazaaatgtg gaggccacta aagagccat  40

<210> SEQ ID NO 81
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 81

tgatgatctgcgtgccatatatgaggagaag

<210> SEQ ID NO: 82
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 82
	tgatgctctctctctcgtgcgtgccatatatgaggagaag

<210> SEQ ID NO: 83
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 83
	asatctctacccctacagctgctagcctaaag

<210> SEQ ID NO: 84
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 84
	cgtgcgtgcgctggtttctgagactaacccttggattgcg

<210> SEQ ID NO: 85
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 85
	gacaggtgcgtgcagcaagagactaacccttggattgcg

<210> SEQ ID NO: 86
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 86
	agtcttcctttctcgtgcagcaagactaacccttggattgcg

<210> SEQ ID NO: 87
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 87
	gacaggtgcgtgcagcaagactaacccttggattgcg
gtcactgtct aaggagctct scgactatctt ggccagagcc
<210> SEQ ID NO 88
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 88
gtgctcaca ccaatatct gcgtggcgc gtcgcgttgc
<210> SEQ ID NO 89
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 89
tgggttctct toccagcaac gtacctgta agatracctc
<210> SEQ ID NO 90
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 90
taaagagttcg tagagcaasct cctgaaaaa agctgagcgc
<210> SEQ ID NO 91
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 91
taagtaagtc ttcagagta tateaaaaa aasgtaestg
<210> SEQ ID NO 92
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 92
taatcaatga gacattacta gocgcagct ttttagaga
<210> SEQ ID NO 93
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 93
gttgactca cactctttta cgaagatct tacggtcaac
<210> SEQ ID NO 94
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 100

```
tttgtaacta atcaacctt tcagccacag gaccacata
```

SEQ ID NO: 101
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 101
```
aaggttctgt cctcataca atacccgaag tcgcgtcat
```

SEQ ID NO: 102
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 102
```
ggaccagcc atcataatac atggtttcat taggcctca
```

SEQ ID NO: 103
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 103
```
cacattgttc actacactgt taaggttggt aagggcctta
```

SEQ ID NO: 104
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 104
```
atacacatt cacccgactgt atttgggccc aaggttttggc
```

SEQ ID NO: 105
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 105
```
cggtctcag gtcgccggtc ttttcagca taaggttaggt
```

SEQ ID NO: 106
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQUENCE: 106
cagcagac ccgctaccag aactaactc gctctcagac

<210> SEQ ID NO: 107
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 107
ctatgggtgt tagcagagt gctcagagt cacccaccc

<210> SEQ ID NO: 108
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 108
cgacagcggc accggagaag tcagagcgt tagcagagc

<210> SEQ ID NO: 109
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 109
cctcggcact ttatggcga aaggacagc gcacagcagc

<210> SEQ ID NO: 110
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 110
gcagcagc ccgtgctcag aatcatacctgt tcaacccasg

<210> SEQ ID NO: 111
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 111
gagatttgct gcagacaaa atgactgaag ggacgttgat

<210> SEQ ID NO: 112
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 112
cagcagac ccctcgtttg tttgattgac ctctca

<210> SEQ ID NO: 113
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 113

gaaagcttct tggcgaagc gcagaaacat gatgacacgc aagc

<210> SEQ ID NO 114
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 114
cgaccattga atagcccaag ggtaatagaga aaccaaggc

<210> SEQ ID NO 115
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 115
gttgaagaaa ggctgataca ccagcagctg ccgccaggg

<210> SEQ ID NO 116
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 116
atactttgag taccacacag aaggtcagaag tgggtgatca

<210> SEQ ID NO 117
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 117
aacgcaacaca gatattcttg tgaatctgca tgcaccttt

<210> SEQ ID NO 118
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 118
agccaaacaca gttgtgcagg aggagcaaa atagccgcc

<210> SEQ ID NO 119
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
-continued

<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 119

actgctcga cgatcgatca ctcgcaagggc ttgaattag

40

<210> SEQ ID NO 120
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 120
cgatatacc atcgctgtaa cgcagatgca atgtagggcg

40

<210> SEQ ID NO 121
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 121
gctctcacag cggagtggat tttcgacggt atccagasta

40

<210> SEQ ID NO 122
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 122
atgatggttt ggtgaaggtt agtacgagac tggacctcca

40

<210> SEQ ID NO 123
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 123
agacotattt caagtgttatt cggcgtagta aaagygggttg

40

<210> SEQ ID NO 124
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 124
aggtttgtcga atacccataa cctcacaag ctcagctctgg

40

<210> SEQ ID NO 125
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 125

atgtaagatt ctttccagg ggccccacgtc atgcgctgtg

<210> SEQ ID NO 126
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 126
accaagcagc gttacagga ataaagagc ggtggtttatt

<210> SEQ ID NO 127
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 127
ctcagcaca atgctcaca cactcttcct tttgtaacca

<210> SEQ ID NO 128
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 128
caatattgga ggtttgggc caccagcaca gtagcttga

<210> SEQ ID NO 129
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 129
aatattttt gtgagggagg ctctcgcttc ccacgagtc

<210> SEQ ID NO 130
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 130
cacgagtgtg tggagggagt gactatggtt acgaggtct

<210> SEQ ID NO 131
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 131
cgaagagca ttcacaccgc ggtgaggtct tccaggtgat

<210> SEQ ID NO 132
<210> SEQ ID NO: 133
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 133
gcacgga tctgtggtcc tgtgtaaat tgttatcgc 40

<210> SEQ ID NO: 134
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 134
ggaacagga tcocatgtg aagctgaga aaatgtcat 40

<210> SEQ ID NO: 135
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 135
catggccct gacgctctcc atcctttgaga ggtttgact 40

<210> SEQ ID NO: 136
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 136
gccggaagcc tgcgtgttcg tgtctcgc aagcacttc 40

<210> SEQ ID NO: 137
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 137
attgctcag accttggtgc gatgtggtcg gogatgaa tc 40
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 138

```
tttgagctaac aagqagttttt tttgagccac gctttgctg
```

<210> SEQ ID NO 139
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 139

```
gctcagtccc tccacaattg tggctacaag atgaagagcg
```

<210> SEQ ID NO 140
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 140

```
tgttagtat cttgctgaa aacaaatacc gttttctcat
```

<210> SEQ ID NO 141
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 141

```
tccagtctac gcggcattgt atctgctatat gctgcggct
```

<210> SEQ ID NO 142
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 142

```
ccagctcaac agagcatact tcctgccgaa ctgtgtaag
```

<210> SEQ ID NO 143
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 143

```
tcaggggtct ttctagcaca cggattgct tccacactaa
```

<210> SEQ ID NO 144
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 144
gaatttctg ascasaagtc sgasagtcsg ascogcacc  

<210> SEQ ID NO 145  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 145  
ascttatta agctatatcat catcttggac actgyggaga  

<210> SEQ ID NO 146  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 146  
statccacgg ttcgastctt ttcgctattc tcatctctcg  

<210> SEQ ID NO 147  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 147  
cattcagag ggcacatacg cacaacttaaa accactccac  

<210> SEQ ID NO 148  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 148  
ttcgacccctg tgcaccaagt tgcagccatt ctggtgacca  

<210> SEQ ID NO 149  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 149  
gcgtactac tggactccca aagggagtca tgcagaccca  

<210> SEQ ID NO 150  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: An oligonucleotide  
<400> SEQUENCE: 150  
tccaaaccctgtcgctgc tgcacattcgctctgcctc  

<210> SEQ ID NO 151
cgctacggc ctcagotgtat tctggtgct acogttctggt 40
<210> SEQ ID NO: 151
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 151
ttcactggcc ttctttccat gottttggtc ttcatatattac 40
<210> SEQ ID NO: 152
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 152
tttcgtttcc tttatggct gccttcggtc gttatgttct 40
<210> SEQ ID NO: 153
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 153
cgccttgggt atcaggggtc tttcttgaaa ggccttccsg 40
<210> SEQ ID NO: 154
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 154
attatggaat cgcagttgtc atcagctgtgc cttaggtgtat 40
<210> SEQ ID NO: 155
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 155
cctgtttttg tttaagagcc caactggtgca caagtctagac 40
<210> SEQ ID NO: 156
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 156
<210> SEQ ID NO: 157
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 157

ttgcttcac tgctgtaatt gtgtgctggt gctgctccac 40

<210> SEQ ID NO 158
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 158
tggctaagga ggtgccgtgaa gtaggcggca acggctgga 40

<210> SEQ ID NO 159
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 159
ttcctcaagg attcttgtaa gttcgctcct ccacgcactc 40

<210> SEQ ID NO 160
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 160
accagcoga ttcttacacg tctctcgagt gsgtticaa 40

<210> SEQ ID NO 161
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 161
gggtcttctt gggccgtgtc actcactca tgyctgctas 40

<210> SEQ ID NO 162
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 162
gatgctgtat cggsgaaactg tgaagcttt gggcctcasa 40

<210> SEQ ID NO 163
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 163
caagtggtgc agctgtgtat cagagccct atggtgagc

<210> SEQ ID NO 164
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 164

aggttatgt ctaaagct gcagctaccaa agagcccat

<210> SEQ ID NO 165
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 165
cagagccggc ggctgtgtgc attcctgtga ttttgtatat

<210> SEQ ID NO 166
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 166
taagcaagc atgagcatat ttacgtctgt gatctttaca

<210> SEQ ID NO 167
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 167

aggagctgt ctaaacaag ggtagccagg ttgctccagc

<210> SEQ ID NO 168
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 168
tgagttggag gagattctgt tgaaaaaatcc atgcattgc

<210> SEQ ID NO 169
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 169
gatgctcagt tggcaggcat tgtctgtctg gaggcgggcg

<210> SEQ ID NO 170
aactgcttc tgccccggt gctcaagcgc ctggtaaaga

attacccgc aaagaagctg atgattactt ggctgaacgt

gtgaagcgata ctaagctactt gcttgccagc gtgcgttttg

ttgacotcct cocctcgtaac gtaacaggca aatattacccg

cagggagctg ttgacacat tgggtgaagaa ggcgcgcgcgt

tagtaaagtgc ttcgtgatta tatagaaaaa aagctagttg
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 176

```
tactcagaa gcctttaact aacggcgccc ttctcaca
```

<210> SEQ ID NO 177
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 177

```
atggtttcag cagctcttg cgggtatatt tgctggttaa
```

<210> SEQ ID NO 178
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 178

```
gtaagaggg atggaatcag caaaagcac gcggccacgc
```

<210> SEQ ID NO 179
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 179

```
aagtaacctag tataagctac aagtttggcc aggtaatctat
```

<210> SEQ ID NO 180
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 180

```
asatctttt ggctttaaat tccttacag gtgtggtgac
```

<210> SEQ ID NO 181
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 181

```
acagccacag gaggccatt ggccggttc cagctcggga
```

<210> SEQ ID NO 182
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 182

```
asgagaaag gaggccagtt ggccggcctc cagctcggga
```
<210> SEQ ID NO: 183
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 183
atgcgcacca cagcgcaccc gcgatgcact ggtttttctc 40

<210> SEQ ID NO: 184
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 184
acggaatctc ctccacactc gcgtgagcfa ccctgctacc 40

<210> SEQ ID NO: 185
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 185
cggagatttg atacgctct tgtaagctc cagcagtaa 40

<210> SEQ ID NO: 186
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 186
aatgcgcacat ctcgctgctga tgcgcacaaa tcacgcagtc 40

<210> SEQ ID NO: 187
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 187
gacagttattg acataacccct tcgcacccat agggcctttg 40

<210> SEQ ID NO: 188
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 188
atgcgcacca cggctaacttgg gttagggccc aagcgtcctac 40

<210> SEQ ID NO: 189
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 189

cagtttcgat atcagcagtc ttacagcaca tgagtggagtt

<210> SEQ ID NO 190
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 190

gacaagccca aaagagcgc tottaaactc atcgcgaga

<210> SEQ ID NO 191
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 191

gactgaaat taggcctggt aagttcgtgtagcgcga

<210> SEQ ID NO 192
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 192

agccacacg aactcctgga aagttcagcg gttgagggct cac

<210> SEQ ID NO 193
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 193

ttcagcagcc ttttagcaca gttgaggagc aagccacacg

<210> SEQ ID NO 194
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 194

aattcagcga gtagagccaa gttgactttg ttccagcagtt

<210> SEQ ID NO 195
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
gacggcagca astgttttga tgggtctgca tgactcctt

<210> SEQ ID NO: 202
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 202

tggagtccg ctagtacgct gctatcagag astggctgca

<210> SEQ ID NO: 203
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 203

acgtgttcga cgggtcgaa gttgaattgt ttaagtgtg

<210> SEQ ID NO: 204
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 204
cgaagtccg cgtctgaatac cgsagatag aastaggcag

<210> SEQ ID NO: 205
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 205

agattcgcgg cggtaaatc tctccacagt gtccagagt

<210> SEQ ID NO: 206
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 206

atgatacgct taaaaggtt ggtcgggctt tggacttccg

<210> SEQ ID NO: 207
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 207

ggacgttgtt cagataatttc ttagtggctg agacaatctg

<210> SEQ ID NO: 208
-continued

<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 209

tggcttgaag ataccaatga ctttacacag ttgctcgga

<210> SEQ ID NO: 209
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 209

atgtagctct gttgactgg agcagcctc atacagatct

<210> SEQ ID NO: 210
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 210

acatgcggc gatgcgtgga atgaagaacctggcttgtt

<210> SEQ ID NO: 211
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 211

ttgacgacag atactaagca cgtggtctat cttgtaggca

<210> SEQ ID NO: 212
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 212

catgtgtga ggacagtgac cagcagacag gtgctcag

<210> SEQ ID NO: 213
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 213

aaaactcttt gtagacaaa gattcactgc gcaccacactc

<210> SEQ ID NO: 214
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQ ID NO 215
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQ ID NO 216
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: An oligonucleotide

SEQ ID NO 217
LENGTH: 542
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Sequence of a synthetic luciferase

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5    10    15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
20   25   30
Lys His Ser His Leu Pro Glu Ala Leu Val Val Gly Asp Glu
35  40   45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
50  55   60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
65  70    75   80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr
115 120 125
<table>
<thead>
<tr>
<th>Lys</th>
<th>Asn</th>
<th>Ile</th>
<th>Leu</th>
<th>Asn</th>
<th>Lys</th>
<th>Val</th>
<th>Leu</th>
<th>Glu</th>
<th>Val</th>
<th>Gln</th>
<th>Ser</th>
<th>Arg</th>
<th>Thr</th>
<th>Asn</th>
<th>Phe</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Lys</td>
<td>Arg</td>
<td>Ile</td>
<td>Ile</td>
<td>Leu</td>
<td>Asp</td>
<td>Thr</td>
<td>Val</td>
<td>Glu</td>
<td>Asn</td>
<td>Ile</td>
<td>His</td>
<td>Gly</td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Ser</td>
<td>Leu</td>
<td>Pro</td>
<td>Asn</td>
<td>Phe</td>
<td>Ile</td>
<td>Ser</td>
<td>Arg</td>
<td>Tyr</td>
<td>Ser</td>
<td>Asp</td>
<td>Gly</td>
<td>Asn</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>175</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Phe</td>
<td>Lys</td>
<td>Pro</td>
<td>Leu</td>
<td>His</td>
<td>Phe</td>
<td>Asp</td>
<td>Pro</td>
<td>Val</td>
<td>Glu</td>
<td>Gln</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>190</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Cys</td>
<td>Ser</td>
<td>Ser</td>
<td>Gly</td>
<td>Thr</td>
<td>Thr</td>
<td>Gly</td>
<td>Leu</td>
<td>Pro</td>
<td>Lys</td>
<td>Gly</td>
<td>Val</td>
<td>Met</td>
<td>Gln</td>
<td>Thr</td>
</tr>
<tr>
<td>195</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Gln</td>
<td>Asn</td>
<td>Ile</td>
<td>Cys</td>
<td>Val</td>
<td>Arg</td>
<td>Leu</td>
<td>Ile</td>
<td>His</td>
<td>Ala</td>
<td>Leu</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
<td>Tyr</td>
</tr>
<tr>
<td>210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Thr</td>
<td>Gln</td>
<td>Leu</td>
<td>Ile</td>
<td>Pro</td>
<td>Gly</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
<td>Tyr</td>
<td>Leu</td>
<td>Pro</td>
<td>Phe</td>
</tr>
<tr>
<td>225</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>His</td>
<td>Ala</td>
<td>Phe</td>
<td>Gly</td>
<td>Phe</td>
<td>His</td>
<td>Ile</td>
<td>Thr</td>
<td>Leu</td>
<td>Gly</td>
<td>Tyr</td>
<td>Phe</td>
<td>Met</td>
<td>Val</td>
<td>Gly</td>
</tr>
<tr>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Arg</td>
<td>Val</td>
<td>Ile</td>
<td>Met</td>
<td>Phe</td>
<td>Arg</td>
<td>Arg</td>
<td>Phe</td>
<td>Asp</td>
<td>Gln</td>
<td>Glu</td>
<td>Ala</td>
<td>Phe</td>
<td>Leu</td>
<td>Lys</td>
</tr>
<tr>
<td>260</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ile</td>
<td>Gln</td>
<td>Asp</td>
<td>Tyr</td>
<td>Glu</td>
<td>Val</td>
<td>Pro</td>
<td>Ser</td>
<td>Val</td>
<td>Ile</td>
<td>Asn</td>
<td>Val</td>
<td>Pro</td>
<td>Ser</td>
<td>Val</td>
</tr>
<tr>
<td>275</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Ser</td>
<td>Lys</td>
<td>Ser</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td>Asp</td>
<td>Lys</td>
<td>Tyr</td>
<td>Asp</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td>290</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Leu</td>
<td>Arg</td>
<td>Glu</td>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
<td>Ala</td>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Val</td>
<td>Ala</td>
<td>Ala</td>
<td>Lys</td>
<td>Arg</td>
<td>Leu</td>
<td>Asn</td>
<td>Leu</td>
<td>Pro</td>
<td>Gly</td>
<td>Ile</td>
<td>Arg</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Gly</td>
<td>Leu</td>
<td>Thr</td>
<td>Glu</td>
<td>Ser</td>
<td>Thr</td>
<td>Ser</td>
<td>Ala</td>
<td>Ile</td>
<td>Ile</td>
<td>Gln</td>
<td>Ser</td>
<td>Leu</td>
<td>Arg</td>
<td>Asp</td>
</tr>
<tr>
<td>340</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Phe</td>
<td>Lys</td>
<td>Ser</td>
<td>Gly</td>
<td>Ser</td>
<td>Leu</td>
<td>Gly</td>
<td>Arg</td>
<td>Val</td>
<td>Thr</td>
<td>Pro</td>
<td>Leu</td>
<td>Met</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>355</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Ile</td>
<td>Ala</td>
<td>Asp</td>
<td>Arg</td>
<td>Glu</td>
<td>Thr</td>
<td>Gly</td>
<td>Ala</td>
<td>Leu</td>
<td>Gly</td>
<td>Pro</td>
<td>Asn</td>
<td>Gln</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>370</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Glu</td>
<td>Leu</td>
<td>Cys</td>
<td>Ile</td>
<td>Lys</td>
<td>Gly</td>
<td>Pro</td>
<td>Met</td>
<td>Val</td>
<td>Ser</td>
<td>Lys</td>
<td>Gly</td>
<td>Tyr</td>
<td>Val</td>
<td>Asn</td>
</tr>
<tr>
<td>385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
<td>Thr</td>
<td>Lys</td>
<td>Glu</td>
<td>Ala</td>
<td>Ile</td>
<td>Asp</td>
<td>Asp</td>
<td>Asp</td>
<td>Gly</td>
<td>Trp</td>
<td>Leu</td>
<td>His</td>
</tr>
<tr>
<td>405</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Gly</td>
<td>Asp</td>
<td>Phe</td>
<td>Gly</td>
<td>Tyr</td>
<td>Tyr</td>
<td>Asp</td>
<td>Glu</td>
<td>Asp</td>
<td>Glu</td>
<td>His</td>
<td>Phe</td>
<td>Tyr</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>420</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Arg</td>
<td>Tyr</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Ile</td>
<td>Lys</td>
<td>Tyr</td>
<td>Lys</td>
<td>Gly</td>
<td>Ser</td>
<td>Glu</td>
<td>Val</td>
<td>Ala</td>
<td>Pro</td>
</tr>
<tr>
<td>435</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Glu</td>
<td>Leu</td>
<td>Glu</td>
<td>Ile</td>
<td>Leu</td>
<td>Leu</td>
<td>Lys</td>
<td>Asn</td>
<td>Pro</td>
<td>Cys</td>
<td>Ile</td>
<td>Arg</td>
<td>Asp</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Val</td>
<td>Val</td>
<td>Gly</td>
<td>Ile</td>
<td>Pro</td>
<td>Asp</td>
<td>Leu</td>
<td>Glu</td>
<td>Ala</td>
<td>Gly</td>
<td>Glu</td>
<td>Leu</td>
<td>Pro</td>
<td>Ser</td>
<td>Ala</td>
</tr>
<tr>
<td>465</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>Val</td>
<td>Val</td>
<td>Lys</td>
<td>Gln</td>
<td>Pro</td>
<td>Gly</td>
<td>Lys</td>
<td>Glu</td>
<td>Ile</td>
<td>Thr</td>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
<td>Val</td>
<td>Tyr</td>
</tr>
<tr>
<td>485</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Tyr</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Arg</td>
<td>Val</td>
<td>Ser</td>
<td>His</td>
<td>Thr</td>
<td>Lys</td>
<td>Tyr</td>
<td>Leu</td>
<td>Arg</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Arg</td>
<td>Phe</td>
<td>Val</td>
<td>Asp</td>
<td>Ser</td>
<td>Ile</td>
<td>Pro</td>
<td>Arg</td>
<td>Asn</td>
<td>Val</td>
<td>Thr</td>
<td>Gly</td>
<td>Lys</td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>515</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Lys</td>
<td>Glu</td>
<td>Leu</td>
<td>Leu</td>
<td>Lys</td>
<td>Gln</td>
<td>Leu</td>
<td>Leu</td>
<td>Glu</td>
<td>Lys</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td>Gly</td>
<td>Gly</td>
</tr>
<tr>
<td>525</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
-continued

530 535 540

<210> SEQ ID NO 219
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 219

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
20  25  30
Lys His Ser His Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu
35  40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
50  55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
65  70  75  80
Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr
115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gln Ser Arg Thr Asn Phe
130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys
145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165 170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Gln Val Ala Ala Ile
180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
195 200 205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr
210 215 220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
225 230 235 240
Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly
245 250 255
Leu Arg Val Ile Met Phe Arg Phe Asp Gln Glu Ala Phe Leu Lys
260 265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
290 295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305 310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Ser Leu Arg Asp
...-continued

340  345  350
Glu Phe Lys Ser Gly Ser Leu Gly Gly Val Thr Pro Leu Met Ala Ala 355  360  365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Gly Pro Asn Gln Val 370  375  380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn 385  390  395  400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Asp Gly Trp Leu His 405  410  415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu His Phe Tyr Val Val 420  425  430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro 435  440  445
Ala Glu Leu Glu Ile Leu Lys Ala Gly Val Leu Pro Cys Ile Arg Asp Val 450  455  460
Ala Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala 465  470  475  480
Phc Val Val Lys Glu Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr 485  490  495
Asp Tyr Leu Ala Gly Asp Val Ser His Thr Lys Tyr Leu Arg Asp Gly Gly 500  505  510
Val Arg Phe Val Asp Ser Ile Pro Arg Asp Val Thr Gly Lys Ile Thr 515  520  525
Arg Lys Glu Leu Leu Lys Glu Leu Leu Glu Lys Ala Gly Gly 530  535  540

<210> SEQ ID NO: 220
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE: -
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 220

Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His 1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg 20  25  30
Lys His Ser Tyr Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu 35  40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gin 50  55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys 65  70  75  80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr 85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu 100  105  110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr 115  120  125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe 130  135  140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165
170
175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile
180
185
190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
195
200
205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr
210
215
220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
225
230
235
240
Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly
245
250
255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys
260
265
270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275
280
285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Tyr Asp Leu Ser
290
295
300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305
310
315
320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325
330
335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gin Ser Leu Arg Asp
340
345
350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
355
360
365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gin Val
370
375
380
Gly Glu Leu Cys Ile Gly Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385
390
395
400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Thr Pro Leu His
405
410
415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val
420
425
430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Gly Ser Gin Val Ala Pro
435
440
445
Ala Glu Leu Glu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
450
455
460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
465
470
475
480
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485
490
495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500
505
510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515
520
525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly Gly
530
535
540

<210> SEQ ID NO 221
<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met</td>
<td>Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His</td>
</tr>
<tr>
<td></td>
<td>1 5 10 15</td>
</tr>
<tr>
<td>Pro</td>
<td>Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg</td>
</tr>
<tr>
<td></td>
<td>20 25 30</td>
</tr>
<tr>
<td>Lys</td>
<td>His Ser His Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu</td>
</tr>
<tr>
<td></td>
<td>35 40 45</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln</td>
</tr>
<tr>
<td></td>
<td>50 55 60</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys</td>
</tr>
<tr>
<td></td>
<td>65 70 75 80</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr</td>
</tr>
<tr>
<td></td>
<td>85 90 95</td>
</tr>
<tr>
<td>Ile</td>
<td>Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu</td>
</tr>
<tr>
<td></td>
<td>100 105 110</td>
</tr>
<tr>
<td>Leu</td>
<td>Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr</td>
</tr>
<tr>
<td></td>
<td>115 120 125</td>
</tr>
<tr>
<td>Lys</td>
<td>Asn Ile Leu Asn Lys Val Leu Glu Val Glu Ser Arg Thr Asn Phe</td>
</tr>
<tr>
<td></td>
<td>130 135 140</td>
</tr>
<tr>
<td>Ile</td>
<td>Lys Arg Ile Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys</td>
</tr>
<tr>
<td></td>
<td>145 150 155 160</td>
</tr>
<tr>
<td>Glu</td>
<td>Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala</td>
</tr>
<tr>
<td></td>
<td>165 170 175</td>
</tr>
<tr>
<td>Asn</td>
<td>Phe Lys Pro Leu His Phe Asp Pro Val Gln Val Ala Ala Ile</td>
</tr>
<tr>
<td></td>
<td>180 185 190</td>
</tr>
<tr>
<td>Leu</td>
<td>Cys Ser Ser Gly Thr Thr Leu Pro Lys Gly Val Met Glu Thr</td>
</tr>
<tr>
<td></td>
<td>195 200 205</td>
</tr>
<tr>
<td>His</td>
<td>Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr</td>
</tr>
<tr>
<td></td>
<td>210 215 220</td>
</tr>
<tr>
<td>Gly</td>
<td>Thr Gln Leu Ile Pro Gly Val Val Val Val Tyr Leu Pro Phe</td>
</tr>
<tr>
<td></td>
<td>225 230 235 240</td>
</tr>
<tr>
<td>Phe</td>
<td>His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly</td>
</tr>
<tr>
<td></td>
<td>245 250 255</td>
</tr>
<tr>
<td>Leu</td>
<td>Arg Val Ile Met Phe Arg Phe Asp Gln Glu Ala Phe Leu Lys</td>
</tr>
<tr>
<td></td>
<td>260 265 270</td>
</tr>
<tr>
<td>Ala</td>
<td>Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val</td>
</tr>
<tr>
<td></td>
<td>275 280 285</td>
</tr>
<tr>
<td>Ile</td>
<td>Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Thr Asp Leu Ser</td>
</tr>
<tr>
<td></td>
<td>290 295 300</td>
</tr>
<tr>
<td>Ser</td>
<td>Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val</td>
</tr>
<tr>
<td></td>
<td>305 310 315 320</td>
</tr>
<tr>
<td>Ala</td>
<td>Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly</td>
</tr>
<tr>
<td></td>
<td>325 330 335</td>
</tr>
<tr>
<td>Phe</td>
<td>Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Ser Leu Arg Asp</td>
</tr>
<tr>
<td></td>
<td>340 345 350</td>
</tr>
<tr>
<td>Glu</td>
<td>Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala</td>
</tr>
<tr>
<td></td>
<td>355 360 365</td>
</tr>
</tbody>
</table>
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
370  375  380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385  390  395  400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
405  410  415
Ser Gly Asp Phe Gly Tyr Asp Glu Asp His Phe Tyr Val Val
420  425  430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro
435  440  445
Ala Glu Leu Glu Ile Leu Lys Asn Pro Cys Ile Arg Asp Val
450  455  460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Leu Pro Ser Ala
465  470  475  480
Phe Val Val Lys Glu Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485  490  495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500  505  510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515  520  525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly Gly
530  535  540

&lt;210&gt; SEQ ID NO 222
&lt;211&gt; LENGTH: 542
&lt;212&gt; TYPE: PRT
&lt;213&gt; ORGANISM: Artificial Sequence
&lt;220&gt; FEATURE:
&lt;222&gt; OTHER INFORMATION: Sequence of a synthetic luciferase
&lt;400&gt; SEQUENCE: 222
Met Met Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg
20  25  30
Lys His Ser His Leu Pro Gin Ala Leu Val Asp Val Val Gly Asp Glu
35  40  45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Glu
50  55  60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys
65  70  75  80
Ala Glu Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr
85  90  95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu
100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gin Ile Val Phe Thr Thr
115 120 125
Leu Asn Ile Leu Asn Gly Val Leu Glu Gin Ser Arg Thr Asn Phe
130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Gly Asn Asn His Gly Cys
145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala
165 170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gln Val Ala Ala Ile
180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gln Thr
195 200 205
His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr
210 215 220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
225 230 235 240
Phe His Ala Phe Gly Phe His Ile Thr Leu G1y Tyr Phe Met Val Gly
245 250 255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys
260 265 270
Ala Ile Gin Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
290 295 300
Ser Leu Arg Glu Leu Cys Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305 310 315 320
Ala Glu Val Ala Ala Gin Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gin Ser Leu Gly Asp
340 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gin Val
370 375 380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Asp Gly Trp Leu His
405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp His Phe Tyr Val Val
420 425 430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gin Ser Glu Val Ala Pro
435 440 445
Ala Glu Leu Glu Glu Ile Leu Leu Asn Pro Cys Ile Arg Asp Val
450 455 460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala
465 470 475 480
Phe Val Val Lys Gin Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485 490 495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500 505 510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515 520 525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Ala Gly Gly
530 535 540

<210> SEQ ID NO 223
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 223

Met Ile Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His 1 5 10 15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg 20 25 30
Lys His Ser His Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu 35 40 45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln 50 55 60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys 65 70 75 80
Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr 85 90 95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu 100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Glu Ile Val Phe Thr Thr 115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gln Ser Arg Thr Asn Phe 130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys 145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala 165 170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gln Val Ala Ala Ile 180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Glu Thr 195 200 205
His Glu Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr 210 215 220
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe 225 230 235 240
Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly 245 250 255 260
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Glu Ala Phe Leu Lys 265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val 275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Tyr Asp Leu Ser 290 295 300
Ser Leu Arg Glu Leu Cys Gly Ala Ala Pro Leu Ala Lys Glu Val 305 310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly 325 330 335 340
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Thr Leu Gly Asp 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala 355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val 370 375 380
| Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn | 385 390 395 400 |
| Aasn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His | 405 410 415 |
| Ser Gly Asp Phe Gly Tyr Asp Gly Asp Glu His Phe Tyr Val Val | 420 425 430 |
| Asp Arg Tyr Lys Glu Ile Lys Tyr Lys Gly Ser Glu Val Ala Pro | 435 440 445 |
| Ala Glu Leu Glu Ile Leu Gln Pro Cys Ile Arg Asp Val | 450 455 460 |
| Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Glu Leu Pro Ser Ala | 465 470 475 480 |
| Phe Val Lys Gln Pro Gly Thr Glu Ile Thr Ala Lys Glu Val Tyr | 485 490 495 |
| Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly | 500 505 510 |
| Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr | 515 520 525 |
| Arg Lys Glu Leu Leu Lys Gin Leu Leu Val Lys Ala Gly Gly | 530 535 540 |

<210> SEQ ID NO 224
<211> LENGTH: 311
<212> TYPE: PRT
<213> ORGANISM: Renilla reniformis

<400> SEQUENCE: 224

Met Thr Ser Lys Val Tyr Asp Pro Glu Gln Arg Lys Arg Met Ile Thr 1   5   10   15
Gly Pro Gln Trp Trp Ala Arg Cys Lys Gin Met Asn Val Leu Asp Ser 20  25  30
Phe Ile Asn Tyr Asp Ser Glu Lys His Ala Glu Asn Ala Val Ile 35  40  45
Phe Leu His Gly Asn Ala Ala Ser Ser Tyr Leu Trp Arg His Val 50  55  60
Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp Leu Ile Gly 65  70  75  80
Met Gly Lys Ser Gly Lys Ser Gly Asn Gly Ser Tyr Arg Leu Leu Asp 85  90  95
His Tyr Lys Tyr Leu Thr Ala Trp Phe Glu Leu Leu Asn Leu Pro Lys 100 105 110
Lys Ile Ile Phe Val Gly His Asp Trp Gly Ala Cys Leu Ala Phe His 115 120 125
Tyr Ser Tyr Glu His Gln Asp Lys Ala Ile Val His Ala Glu 130 135 140
Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro Asp Ile Glu 145 150 155 160
Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Lys Met Val Leu 165 170 175
Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys Ile Met Arg 180 185 190
Lys Leu Glu Pro Glu Glu Phe Ala Ala Tyr Leu Glu Pro Phe Lys Glu 195 200 205
Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg Glu Ile Pro
210     215     220

Leu Val Lys Gly Gly Lys Pro Asp Val Val Gin Ile Val Arg Asn Tyr
225     230     235     240

Asn Ala Tyr Leu Arg Ala Ser Asp Leu Pro Lys Met Phe Ile Glu
245     250     255

Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly Ala Lys Lys
260     265     270

Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His Phe Ser Gln
275     280     285

Glu Asp Ala Pro Asp Glu Met Gly Lys Tyr Ile Lys Ser Phe Val Glu
290     295     300

Arg Val Leu Lys Asn Glu Gln
305     310

<210> SEQ ID NO 225
<211> LENGTH: 311
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 225

Met Ala Ser Lys Val Tyr Asp Pro Glu Gin Arg Lys Arg Met Ile Thr
  1     5      10     15
Gly Pro Gln Thr Thr Ala Arg Cys Lys Gin Met Asn Val Leu Asp Ser
 20     25     30     35
Phe Ile Asn Tyr Tyr Asp Ser Glu Lys His Ala Glu Asn Ala Val Ile
  35     40     45     50
Phe Leu His Gly Asn Ala Ser Ser Tyr Leu Thr Trp Arg His Val Val
  55     60
Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp Leu Ile Gly
  65     70     75     80
Met Gly Lys Ser Gly Ser Gly Aan Gly Ser Tyr Arg Leu Asp
  85     90     95
His Tyr Lys Tyr Leu Thr Ala Thr Phe Glu Leu Leu Asn Leu Pro Lys
100    105    110
Lys Ile Ile Phe Val Gly His Asp Trp Gly Ala Cys Leu Ala Phe His
115    120    125
Tyr Ser Tyr Glu His Gin Asp Lys Ile Lys Ala Ile Val His Ala Glu
130    135    140
Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro Asp Ile Glu
145    150    155    160
Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Glu Lys Met Val Leu
165    170    175
Glu Asn Gin Phe Phe Val Glu Thr Met Leu Pro Ser Lys Ile Met Arg
180    185    190
Lys Leu Glu Pro Glu Glu Phe Ala Tyr Leu Glu Pro Phe Lys Glu
195    200    205
Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg Glu Ile Pro
210    215    220
Leu Val Lys Gly Gly Pro Asp Val Val Gin Ile Val Arg Asn Tyr
225    230    235    240
Asn Ala Tyr Leu Arg Ala Ser Asp Asp Leu Pro Lys Met Phe Ile Glu
245  250  255
Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly Ala Lys Lys
260  265  270
Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His Phe Ser Glu
275  280  285
Glu Asp Ala Pro Asp Glu Met Gly Lys Tyr Ile Lys Ser Phe Val Glu
290  295  300
Arg Val Leu Lys Asn Glu Gln
305  310

<210> SEQ ID NO 226
<211> LENGTH: 311
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCES: 226

Met Ala Ser Lys Val Tyr Asp Pro Glu Gln Arg Lys Arg Met Ile Thr
1   5   10   15
Gly Pro Gln Trp Trp Ala Arg Cys Lys Gln Met Asn Val Leu Asp Ser
20  25   30
Phe Ile Asn Tyr Tyr Asp Ser Glu His Ala Glu Asn Ala Val Ile
35   40   45
Phe Leu His Gly Asn Ala Ala Ser Ser Tyr Leu Trp Arg His Val Val
50   55   60
Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp Leu Ile Gly
65   70   75   80
Met Gly Lys Ser Gly Lys Ser Gly Asn Gly Ser Tyr Arg Leu Leu Asp
85   90   95
His Tyr Lys Tyr Leu Thr Ala Trp Phe Glu Leu Leu Asn Leu Pro Lys
100  105  110
Lys Ile Ile Phe Val Gly His Asp Trp Gly Ala Cys Leu Ala Phe His
115  120  125
Tyr Ser Tyr Glu His Gln Asp Lys Ile Lys Ala Ile Val His Ala Glu
130  135  140
Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro Asp Ile Glu
145  150  155  160
Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Glu Lys Met Val Leu
165  170  175
Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys Ile Met Arg
180  185  190
Lys Leu Glu Pro Glu Glu Phe Ala Tyr Leu Glu Pro Phe Lys Glu
195  200  205
Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg Glu Ile Pro
210  215  220
Leu Val Lys Gly Gly Lys Pro Asp Val Val Glu Ile Val Arg Asn Tyr
225  230  235  240
Asn Ala Tyr Leu Arg Ala Ser Asp Asp Leu Pro Met Phe Ile Glu
245  250  255
Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly Ala Lys Lys
260  265  270
Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His Phe Ser Gln 275 280 285
Glu Asp Ala Pro Asp Glu Met Gly Tyr Ile Lys Ser Phe Val Glu 290 295 300
Arg Val Leu Lys Asn Glu Gin 305 310

<210> SEQ ID NO 227
<211> LENGTH: 311
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 227
Met Ala Ser Lys Val Tyr Asp Pro Glu Gin Arg Lys Arg Met Ile Thr 1 5 10 15
Gly Pro Gln Trp Trp Ala Arg Cys Lys Gin Met Asn Val Leu Asp Ser 20 25 30
Phe Ile Asn Tyr Tyr Asp Ser Glu His Ala Glu Asn Ala Val Ile 35 40 45
Phe Leu His Gly Asn Ala Ser Ser Tyr Leu Trp Arg His Val Val 50 55 60
Pro His Ile Glu Pro Val Ala Arg Cys Ile Ile Pro Asp Leu Ile Gly 65 70 75 80
Met Gly Lys Ser Gly Lys Ser Gly Arg Asp Leu Gin Gin 85 90 95
His Tyr Lys Tyr Leu Thr Ala Trp Phe Glu Leu Leu Asn Leu Pro Lys 100 105 110
Lys Ile Ile Phe Val Gly His Asp Trp Gly Ala Cys Leu Ala Phe His 115 120 125
Tyr Ser Tyr Glu His Gin Asp Lys Ile Lys Ala Ile Val His Ala Glu 130 135 140
Ser Val Val Asp Val Ile Glu Ser Trp Asp Glu Trp Pro Asp Ile Glu 145 150 155 160
Glu Asp Ile Ala Leu Ile Lys Ser Glu Glu Gly Lys Met Val Leu 165 170 175
Glu Asn Asn Phe Phe Val Glu Thr Met Leu Pro Ser Lys Ile Met Arg 180 185 190
Lys Leu Glu Pro Glu Glu Phe Ala Ala Tyr Leu Glu Pro Phe Lys Glu 195 200 205
Lys Gly Glu Val Arg Arg Pro Thr Leu Ser Trp Pro Arg Glu Ile Pro 210 215 220
Leu Val Lys Gly Gly Lys Pro Asp Val Val Gin Ile Val Arg Asn Tyr 225 230 235 240
Asn Ala Tyr Leu Arg Ala Ser Asp Leu Pro Lys Met Phe Ile Glu 245 250 255
Ser Asp Pro Gly Phe Phe Ser Asn Ala Ile Val Glu Gly Ala Lys Lys 260 265 270
Phe Pro Asn Thr Glu Phe Val Lys Val Lys Gly Leu His Phe Ser Glu 275 280 285
Glu Asp Ala Pro Asp Glu Met Gly Lys Tyr Ile Lys Ser Phe Val Glu 290 295 300
Arg Val Leu Lys Asn Glu Gln
305 310

<210> SEQ ID NO 228
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(14)
<223> OTHER INFORMATION: n = A, T, C or G
<400> SEQUENCE: 228

yggmnmmng ccua 14

<210> SEQ ID NO 229
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 229
gtactgagcc gaagcgcgcc caagcttagg cctgagtg 38

<210> SEQ ID NO 230
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 230
ggcagcagc gtaactgact gaactagcg ccgcccag 38

<210> SEQ ID NO 231
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 231
ggatccatg gtaagcgtg agaa 24

<210> SEQ ID NO 232
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 232
ggatccatg gtaagcctg a 21

<210> SEQ ID NO 233
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 233
ctagtttt tttctagata atcatgaaga c
                        31

<210> SEQ ID NO: 234
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 234
caaaaagctt gccattccgg tactgtttgtg aagcccacca ttgtagaagc agag
                        54

<210> SEQ ID NO: 235
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 235
cattgtgt tgttaaccttg tttatt
                        26

<210> SEQ ID NO: 236
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 236
acacattggt tcaaggtgt aagaccccga gcaacgaaa
                        40

<210> SEQ ID NO: 237
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 237
gccttggagct tactgctcgct tottccagcgc gcgcctcagy
                        40

<210> SEQ ID NO: 238
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 238
cgcctcgcct gcgttggaga gtttatagtc c
                        31

<210> SEQ ID NO: 239
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 239
ggcccgttac tcctagatttt tgttt
                        25
<210> SEQ ID NO 240
<211> LENGTH: 5
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 240

tatca

5

<210> SEQ ID NO 241
<211> LENGTH: 6
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 241

strtg

6

<210> SEQ ID NO 242
<211> LENGTH: 9
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(9)
<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 242

mtttnnnnma

9

<210> SEQ ID NO 243
<211> LENGTH: 5
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 243

tratg

5

<210> SEQ ID NO 244
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A consensus sequence

<400> SEQUENCE: 244

tgastma

7

<210> SEQ ID NO 245
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A consensus sequence
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(14)
<223> OTHER INFORMATION: n = A,T,C or G
<400> SEQUENCE: 245

ygggannng ccaa 14

<210> SEQ ID NO: 246
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 246

aaccatgtgct tcaccgggtg aacaccacca gcaacgcaaa 40

<210> SEQ ID NO: 247
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 247

cgctgtatca ctgggcctca gtggtgggtg ctgtgcaagc 40

<210> SEQ ID NO: 248
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 248

aaatgaacgt gtggactcc ttcatcaact actagtatc 40

<210> SEQ ID NO: 249
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 249

cgagaacgcc ggcgagaacgg cgggtatttt tctgcatggt aacgtgctct 50

<210> SEQ ID NO: 250
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 250

cagcgatact ctggagcgoc gtcgtgcttc acatcggccc 40

<210> SEQ ID NO: 251
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 251

cggtgctaga tcgtctatcc ctgatctgat cggatggtg 40
<210> SEQ ID NO 252
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 252

aagtcggca agacgcggga tggcctat atlccctctg

<210> SEQ ID NO 253
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 253

atcactaca gtcctccacc gcgtgttgc agctgctgaa

<210> SEQ ID NO 254
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 254

cctttcaasg asatcattct ttgtgggcca cgaatggggg

<210> SEQ ID NO 255
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 255

gctgtgttg cctttcata tctctacag cccacaagca

<210> SEQ ID NO 256
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 256

agatacaggc cagtagctcat gctgagagtg tcttgtgacgt

<210> SEQ ID NO 257
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 257

gatcgagtcg tgggacgagc gcctgacat ogaggaggt atcgc
cctgataaag aagcgaaggg gcggagaaat gggtgcttggag

ataaacttct tcgcgagac caggtctcga agcagatca

tgcggaact ggcgccttgag gagttccttg ccctacctgga gcocat

tcaagagaga ggcgcaggtt agagggccca cccctccttg

gctctgcag atccctctcg ttaagggagg caagcgcagac

gctctgcaga tgcgcgcaca ctcaacgccc taccttcggg
<400> SEQUENCE: 264
cagagagaga tttgtcaag atgttcag agtcgaccc  40

<210> SEQ ID NO: 265
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 265
ttgtcttttt tcaacgctca tgtcgacgg agcaagaag  40

<210> SEQ ID NO: 266
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 266
ttcocctaca cgsagttcgt gaaagtgac ggcctccact  40

<210> SEQ ID NO: 267
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 267
tcagocagga ggcgttctca gatgaaatg gtaagttat  40

<210> SEQ ID NO: 268
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 268
cagagttcct tggagcgcg tgcctgacaa cgacgtaa tcttagacgc  49

<210> SEQ ID NO: 269
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 269
gttcctact atctgtctct tttcagac  29

<210> SEQ ID NO: 270
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 270
cagcgtcggc gaaagttcct tctagcttcac ctatctgc  40
<210> SEQ ID NO 271
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 271
tgaggctcc tctggtctga agtgaggccc cttcacttc 40

<210> SEQ ID NO 272
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 272
acgaacctgg tgtaggagaa cttcttagct ccctcgacca 40

<210> SEQ ID NO 273
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 273
tagctttgga aasgacccca ggctcgact cgatgacac 40

<210> SEQ ID NO 274
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 274
cctaggccga tgtgccttg ccgccaggtg ggcgtttgag 40

<210> SEQ ID NO 275
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 275
ttgccgacca tctggacgc gtcggtcttg cttcccttaa 40

<210> SEQ ID NO 276
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 276
cgagaaggtg ttctggaggg caagagagg taggctgtct 40

<210> SEQ ID NO 277
<211> LENGTH: 40
<212> TYPE: DNA
-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 277

aactgcgccc ttcctcttga atggctcag gtaccagcg

<210> SEQ ID NO 278
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 278

aactctcag gtccagtgtt cagcagatc ttcgcttggga gcag

<210> SEQ ID NO 279
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 279

gctotacga agasgtaatt ctcagcacc attttctgc

<210> SEQ ID NO 280
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 280

cctcttcgct cttgactcgg gogatatcct cctgatgctc

<210> SEQ ID NO 281
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 281

agggcaactg ttcagaagct cagcatcgc tccagacactc tcg

<210> SEQ ID NO 282
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 282

gcatggacga tggctcttgat cttgctctgag tgcctcgagag

<210> SEQ ID NO 283
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide
<400> SEQUENCE: 263

tagtgaaag ccaacaacg cccocagtcg tggccocaaca 40

<210> SEQ ID NO 284
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 284

agatgatattt ctttgaaggg ttccagagct cgaacacagc 40

<210> SEQ ID NO 285
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 285

ggtgaggtac ttgtagttat ccagggccg atatgagcga 40

<210> SEQ ID NO 286
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 286

ttccogctct tgcgagact atccatccg atcagatcag 40

<210> SEQ ID NO 287
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 287

ggatgatga cctagcacaag ggctgatgag gaggacacgt gtgcc 45

<210> SEQ ID NO 288
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 288

tccacaagta gctggagccc gctttacatg ccaaaaaat 40

<210> SEQ ID NO 289
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 289

cagcgagtta ccaggtgtgct ttgaggtact atagtaggtg atgaa 45
<210> SEQ ID NO 290
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 290

ggatccag acgttcattt gcttgcagc agcccaacct  

<210> SEQ ID NO 291
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 291
tgacgccccg gtagcatcgq ttgctttgct cgggtcgt  

<210> SEQ ID NO 292
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 292
acaccttgag agcctggtt  

<210> SEQ ID NO 293
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A Kozak sequence

<400> SEQUENCE: 293
aacccatgct  

<210> SEQ ID NO 294
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: An oligonucleotide

<400> SEQUENCE: 294
taatcttacgc gc  

<210> SEQ ID NO 295
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer

<400> SEQUENCE: 295
gcgatccgac gtagctacgq gcagaasag tc  

<210> SEQ ID NO 296
<211> LENGTH: 33
<212> TYPE: DNA
---continued---

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A primer
<400> SEQUENCE: 296

cgaactctag attactaacc gcgggcttc acc

<210> SEQ ID NO 297
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 297

atggtgaac gcgaaagaac cgtgactac gcgcccagaa cactgcacct acgtgaagac 60
c tacagcctg gtgtgagcct ctctcgacaa ctagctaccc caaacaagaa 120
c tctgtgaggc tctctgagaa ccgagacctc tctctaaaag accctttttaa agctacttgt 180
c tctttggccc aagctcttca tataattgcg tcacaaatgga acgcgttgtt gacctttgtg 240
gctgagaata acacgtcgcct ctttatctct gtatcgtctt ctgctaacct ccgcgtatt 300
gtgggccgct tgtatggtat ttcattcctcta ggtgagctgt gtaaggttat gggatttgcc 360
aacacccaa cttccatc taaaacaaa acctgtaa ctgctatgga agtcacctct 420
cgcactaacc tctactcactat tctctgacgct ccagcagctgcat caagcgctgt 480
gggaccgcct cttcctgctt cttctcgctt acgcgttga attgacactt ggcgtgcgt 540
ttgctattg tgtcagctga gcacagctgc gotatctttgt gcctctcagg caacacttgt 600
tgctctaaag gctctcagtc gacctaccc aatctctgttg tgcgttggct ccacgcctgc 660
gaacgctgg tgggatcctc attgacactt gcgcgtgcgt tgcgtggtgtg tgcgtccttc 720
ccatcggcg tgttcattcatt ctcctcgctt cttcattcgtg ctgctatcgcg gctgtgcttc 780
agtttggcg tgtccagcag acgcgcttc ttcgtctctt ccagcagctgcat ccagcagctg 840
tccgtgatc agtccgcttc acaatatttc gcttctagtc ccagcagctgcat ccagcagctg 900
tattgcttc gcacggcgct ctgcacggcgctg ccacgcttg ccacgcttg ccacgcttg 960
gggccggct tgtcagctta ctctcagctt cctctccttc ggtctgtgctt tcatttccttc 1020
gagagccct cttgaatcct cctctccttc gcgcggtttg cttcatgtgc cggagcactc 1080
cgcgagctcg ctcgcttcgc tgtctcttccc qccaaagagc gcacgctgtg cgcagccttc 1140
cacccaaagctg cttcttcctg gcgcggtttg ccacgcttg ccacgcttg ccacgcttg 1200
acagcttcgtg cttgcagctg cttcttcctg gcgcggtttg ccacgcttg ccacgcttg 1260
<210> SEQ ID NO 298
**LENGTH:** 542  
**TYPE:** PRT  
**ORGANISM:** Artificial Sequence  
**FEATURE:**  
**OTHER INFORMATION:** Sequence of a synthetic luciferase

```plaintext
<400> SEQUENCE: 298

MET VAL LYS ARG GLU LYS ASN VAL ILE TYR GLY PRO GLU PRO LEU HIS
1     5     10
PRO LEU GLU ASP LEU THR ALA GLY GLU MET LEU PHE ARG ALA LEU ARG
20    25    30
LYS HIS SER SER LEU PRO GLN ALA LEU VAL ASP VAL VAL GLY ASP GLU
35    40    45
SER LEU SER TYR LYS GLU PHE PHE GLU ALA THR VAL LEU LEU ALA GLN
50    55    60
SER LEU HIS ASN CYS GLY TYR LYS MET ASN ASP VAL VAL SER ILE CYS
65    70    75    80
ALA GLU ASN ASN THR ARG PHE PHE ILE PRO VAL ILE ALA ALA TRP TYR
85    90    95
ILE GLY MET ILE VAL ALA PRO VAL ASN GLU SER TYR ILE PRO ASP GLU
100   105   110
LEU CYS LYS VAL MET GLY ILE SER LYS PRO GLN ILE VAL PHE THR THR
115   120   125
LYS ASN ILE LEU ASN LYS VAL LEU GLU GLN SER ARG THR ASN PHE
130   135   140
ILE lys arg ile ile leu asp thr val glu asn ile his gly cys
145   150   155   160
GLU SER LEU PRO ASP PHE ILE SER ARG TYR SER ASP GLY ASN ILE ALA
165   170   175
ASN PHE LYS PRO LEU HIS PHE ASP PRO VAL GLU GLN VAL ALA ALA ILE
180   185   190
LEU CYs SER SER GLY THR THR GLU PRO LYS GLY VAL MET GLN THR
195   200   205
HIS GLN ASN ILE CYS VAL ARG LEU ILE HIS ALA LEU ASP PRO ARG VAL
210   215   220
GLY THR GLN LEU ILE PRO GLY VAL THR VAL LEU VAL TYR LEU PRO PHE
225   230   235   240
PHE HIS ALA PHE GLY PHE SER ILE THR LEU GLY TYR PHE MET VAL GLY
245   250   255
LEU ARG VAL ILE MET PHE ARG PHE ASP GLN GLU ALA PHE LEU LYS
260   265   270
ALA ILE GLN ASP TYR GLU VAL ARG SER VAL ILE ASN VAL PRO SER VAL
275   280   285
ILE LEU PHE LEU SER LYS SER PRO LEU VAL ASP TYR ASP LEU SER
290   295   300
SER LEU ARG GLU LEU CYS GLY ALA ALA PRO LEU ALA LYS GLU VAL
305   310   315   320
ALA GLU VAL ALA ALA LYS ARG LEU ASN LEU PRO GLY ILE ARG CYS GLY
325   330   335
PHE GLY LEU THR GLU SER THR SER ALA ASN ILE HIS SER LEU ARG ASP
340   345   350
GLU PHE LYS SER GLY SER LEU GLY ARG VAL THR PRO LEU MET ALA ALA
355   360   365
```
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
370 375 380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Thr Leu His
405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp His Phe Tyr Val Val
420 425 430
Asp Arg Tyr Lys Glu Leu Ile Tyr Lys Gly Ser Gln Val Ala Pro
435 440 445
Ala Glu Leu Glu Glu Ile Leu Lys Asn Pro Cys Ile Arg Asp Val
450 455 460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Leu Pro Ser Ala
465 470 475 480
Phe Val Val Lys Gin Pro Gly Lys Glu Ile Thr Ala Lys Glu Val Tyr
485 490 495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500 505 510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515 520 525
Arg Lys Glu Leu Leu Lys Gin Leu Leu Glu Lys Asa Gly Gly
530 535 540

<210> SEQ ID NO: 299
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<222> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 299

atggtgaaa ggtggaaaaa tttctactat ggcocctgac cttcacaact cttggaggtat 60
ttgtgtcgg cggagaattg tttctgtgct cttgcaacag accttctattt gctctaaagcc 120
ttgctggtg tggctgtgca tgaatctttg agctaacagag aagtttttgag ggcacactgtc 180
ttgtggtgct attctctctc caatggtggc tcaagaagtg aacaagtcatcg tagtatgtgt 240
gtgaacana atacacacct ttcttctcaca gtcacagccg catgtattat catgcatgct 300
gtggtgtagc caaacagca gctacacttg gtaaacagtg ggtctctctt 360
aagcaacca tggctcaacc caatgagatatattgcaac aactctctcg gacgcacactgtc 420
cggcaacaact ttattacagt gattcacttc ttgagactct tggagatatt ctaaagtttgcc 480
gaatctttg cttactcttc atctgagct tggaaacgca acatgaaatt attctcacag 540
cctcaacact ccactgccta atctctctct cttgccagaag ccacggattgt 600
cgctccaggg gactaacgct gacccctgta cagccacctg gctccctcttc 660
gacccactt gcgctc gtttgcctctt ggttttttat gttgtttcaa ccaggtctcc caa 720
tctctctcttt cttgcttatt cttctttctct cttctcttctt ccaggtctcc 780
gagtttcgct gtttgccttt gcagttcttct gttttttcat gttttttttt gttgtttctctt 840
tagctttcagc cagctctgt cttgtctctt cttcatttctt tggctgtctt cttcactcttc 900
ttgattcctt cttctcttcttttctctt tgcgctgtctt tgcgctgtctt gttgtttttttt 960
-continued-

gctgaagttg ccgccaaacgc ccagagccttc gtttgagcctt ccggctcaco 020

gaadacccg cggcatatct cggcattccg ccggatcagc ttaaagaggg ctctttgggc 080

cggatcagc ccactccggtc tgcataagtc ctcgtagcgc ccgctccatg aaactggtaa 140

cgagaacag cggccagcgt gttgtacaa gcggccaggg tcgacaggg ttagctaat 200

aacgatgag ctcagcaagag ggcctacagc gcgcgcgctc gttggcatcct tgggtatttt 260

ggtataaca aacgagagta ccatatctac gcgtgtgattc gcgtgataac ggcctcaca 320

tacgaggtga gcccaggtgtc ttcaggtgag aggttgagga ttagtggaga aacccagagc 380

tacgagtggc ggcgagctct gtagtttggc ggctgaggg ccggcgacac gcctttgcgt 440

tccgtggtca ccggcagcttg ctagaaattc aacgccaaag aaggttagta ttagctgtgt 500

gcagctgta gcctactctg ctagattggt cgcgcgctgc gttgggttga ctcctatccc 560

ggtacagtaa cagggcaaat taccogcaag gagctgtgta aacaattgtt ggagaagccg 620

gccggt 626

<210> SEQ ID NO: 300
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 300

Met Val Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His 1  5  10  15
Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu Phe Arg Ala Leu Arg 20 25 30
Lys His Ser His Leu Pro Gln Ala Leu Val Asp Val Val Gly Asp Glu 35 40 45
Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln 50 55 60
Ser Leu His Asn Cys Gly Tyr Lys Met Asn Asp Val Val Ser Ile Cys 65 70 75 80
Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Trp Tyr 85 90 95
Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu 100 105 110
Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gln Ile Val Phe Thr Thr 115 120 125
Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe 130 135 140
Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys 145 150 155 160
Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala 165 170 175
Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile 180 185 190
Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gin Thr 195 200 205
His Gin Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr 210 215 220
\begin{verbatim}
Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe
225 230 235 240
Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly
245 250 255
Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gln Glu Ala Phe Leu Lys
260 265 270
Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val
275 280 285
Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser
290 295 300
Ser Leu Arg Glu Leu Cys Cys Gly Ala Ala Pro Leu Ala Lys Glu Val
305 310 315 320
Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly
325 330 335
Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gln Ser Leu Arg Asp
340 345 350
Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala
355 360 365
Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gln Val
370 375 380
Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn
385 390 395 400
Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Gly Trp Leu His
405 410 415
Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Asp Glu His Phe Tyr Val Val
420 425 430
Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gln Val Ala Pro
435 440 445
Ala Glu Leu Glu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val
450 455 460
Ala Val Val Gly Ile Pro Asp Leu Glu Ala Gly Leu Pro Ser Ala
465 470 475 480
Phe Val Val Lys Glu Pro Gly Lys Glu Thr Ala Lys Glu Val Tyr
485 490 495
Asp Tyr Leu Ala Glu Arg Val Ser His Thr Lys Tyr Leu Arg Gly Gly
500 505 510
Val Arg Phe Val Asp Ser Ile Pro Arg Asn Val Thr Gly Lys Ile Thr
515 520 525
Arg Lys Glu Leu Leu Lys Glu Leu Glu Aly Ala Gly Gy
530 535 540
<210> SEQ ID NO 301
<211> LENGTH: 1626
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase
<400> SEQUENCE: 301
atggttaaacg tgtsqasaaa tggcatctat gcggcttgac gcctcctacct tttggaggtat
60
ttgaactgcgg gqgaatgct ggctgggt gtctctcgagt ctcctcgatttt gcggacgagc
ttggtcata gttgcggaga tgaatcttttg agctacagsg attttttgg gcggacgagc
ttggtcata tgggtgcagg 180
\end{verbatim}
<210> SEQ ID NO 302
<211> LENGTH: 542
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of a synthetic luciferase

<400> SEQUENCE: 302

Met Val Lys Arg Glu Lys Asn Val Ile Tyr Gly Pro Glu Pro Leu His
1  5  10  15

Pro Leu Glu Asp Leu Thr Ala Gly Glu Met Leu PheArg Ala Leu Arg
20  25  30

Lys His Ser His Leu Pro Gln Ala Leu Val Aep Val Val Gly Asp Glu
35  40  45

Ser Leu Ser Tyr Lys Glu Phe Phe Glu Ala Thr Val Leu Leu Ala Gln
50  55  60

Ser Leu His Asn Cys Gly Tyr Lys Met Asn Aep Val Val Ser Ile Cys
65  70  75  80

Ala Glu Asn Asn Thr Arg Phe Phe Ile Pro Val Ile Ala Ala Ala Trp Tyr
---continued---

<table>
<thead>
<tr>
<th></th>
<th>85</th>
<th>90</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile Gly Met Ile Val Ala Pro Val Asn Glu Ser Tyr Ile Pro Asp Glu</td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Leu Cys Lys Val Met Gly Ile Ser Lys Pro Gin Ile Val Phe Thr Thr</td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Lys Asn Ile Leu Asn Lys Val Leu Glu Val Gin Ser Arg Thr Asn Phe</td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Ile Lys Arg Ile Ile Leu Asp Thr Val Glu Asn Ile His Gly Cys</td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Glu Ser Leu Pro Asn Phe Ile Ser Arg Tyr Ser Asp Gly Asn Ile Ala</td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Asn Phe Lys Pro Leu His Phe Asp Pro Val Glu Gin Val Ala Ala Ile</td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Leu Cys Ser Ser Gly Thr Thr Gly Leu Pro Lys Gly Val Met Gin Thr</td>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>His Gln Asn Ile Cys Val Arg Leu Ile His Ala Leu Asp Pro Arg Tyr</td>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Gly Thr Gln Leu Ile Pro Gly Val Thr Val Leu Val Tyr Leu Pro Phe</td>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Phe His Ala Phe Gly Phe His Ile Thr Leu Gly Tyr Phe Met Val Gly</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Leu Arg Val Ile Met Phe Arg Arg Phe Asp Gin Glu Ala Phe Leu Lys</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Ala Ile Gln Asp Tyr Glu Val Arg Ser Val Ile Asn Val Pro Ser Val</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Ile Leu Phe Leu Ser Lys Ser Pro Leu Val Asp Lys Tyr Asp Leu Ser</td>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Ser Leu Arg Glu Leu Cys Cys Gly Ala Ala Pro Leu Ala Lys Glu Val</td>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Ala Glu Val Ala Ala Lys Arg Leu Asn Leu Pro Gly Ile Arg Cys Gly</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Phe Gly Leu Thr Glu Ser Thr Ser Ala Ile Ile Gin Thr Leu Gly Asp</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Glu Phe Lys Ser Gly Ser Leu Gly Arg Val Thr Pro Leu Met Ala Ala</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Lys Ile Ala Asp Arg Glu Thr Gly Lys Ala Leu Gly Pro Asn Gin Val</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Gly Glu Leu Cys Ile Lys Gly Pro Met Val Ser Lys Gly Tyr Val Asn</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Asn Val Glu Ala Thr Lys Glu Ala Ile Asp Asp Asp Gly Tp Leu His</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Ser Gly Asp Phe Gly Tyr Tyr Asp Glu Gin Glu Gin His Phe Tyr Val Val</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Asp Arg Tyr Lys Glu Leu Ile Lys Tyr Lys Gly Ser Gin Val Ala Pro</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Ala Glu Leu Glu Glu Ile Leu Leu Lys Asn Pro Cys Ile Arg Asp Val</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1-47. (canceled)
48. A method to prepare a synthetic nucleic acid molecule comprising an open reading frame, comprising:

a) altering a plurality of transcription regulatory sequences in a parent nucleic acid sequence which encodes a polypeptide having at least 100 amino acids to yield a synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to the parent nucleic acid sequence, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites, enhancer sequences and promoter sequences; and

b) altering greater than 25% of the codons in the synthetic nucleic acid sequence which has a decreased number of transcription regulatory sequences to yield a further synthetic nucleic acid molecule, wherein the codons which are altered do not result in an increased number of transcription regulatory sequences, wherein the further synthetic nucleic acid molecule encodes a polypeptide with at least 85% amino acid sequence identity to the polypeptide encoded by the parent nucleic acid sequence.

49. A method to prepare a synthetic nucleic acid molecule comprising an open reading frame, comprising:

a) altering greater than 25% of the codons in a parent nucleic acid sequence which encodes a polypeptide having at least 100 amino acids to yield a codon-altered synthetic nucleic acid molecule, and

b) altering a plurality of transcription regulatory sequences in the codon-altered synthetic nucleic acid molecule to yield a further synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to a synthetic nucleic acid molecule with a random selection of codons at the codons which differ, wherein the transcription regulatory sequences are selected from the group consisting of transcription factor binding sequences, intron splice sites, poly(A) addition sites, enhancer sequences and promoter sequences, and wherein the further synthetic nucleic acid molecule encodes a polypeptide with at least 85% amino acid sequence identity to the polypeptide encoded by the parent nucleic acid sequence.

50. The method of claim 48 or 49 wherein the parent nucleic acid sequence encodes a reporter molecule.

51. The method of claim 48 or 49 wherein the parent nucleic acid sequence encodes a luciferase.

52. The method of claim 48 or 49 wherein the synthetic nucleic acid molecule hybridizes under medium stringency hybridization conditions to the parent nucleic acid sequence.

53. The method of claim 48 or 49 wherein the codons which are altered encode the same amino acid as the corresponding codons in the parent nucleic acid sequence.

54. (canceled)

55. A method for preparing at least two synthetic nucleic acid molecules which are codon distinct versions of a parent nucleic acid sequence which encodes a polypeptide, comprising:

a) altering a parent nucleic acid sequence to yield a synthetic nucleic acid molecule having an increased number of a first plurality of codons that are employed more frequently in a selected host cell relative to the number of those codons in the parent nucleic acid sequence; and

b) altering the parent nucleic acid sequence to yield a further synthetic nucleic acid molecule having an increased number of a second plurality of codons that are employed more frequently in the host cell relative to the number of those codons in the parent nucleic acid sequence, wherein the first plurality of codons is different than the second plurality of codons, and wherein the synthetic and the further synthetic nucleic acid molecules encode the same polypeptide.

56. The method of claim 55 further comprising altering a plurality of transcription regulatory sequences in the synthetic nucleic acid molecule, the further synthetic nucleic acid molecule, or both, to yield at least one yet further synthetic nucleic acid molecule which has at least 3-fold fewer transcription regulatory sequences relative to the synthetic nucleic acid molecule, the further synthetic nucleic acid molecule, or both.

57. The method of claim 55 further comprising altering at least one codon in the first synthetic sequence to yield a first modified synthetic sequence which encodes a polypeptide with at least one amino acid substitution relative to the polypeptide encoded by the first synthetic nucleic acid sequence.

58. The method of claim 56 further comprising altering at least one codon in the second synthetic sequence to yield a second modified synthetic sequence which encodes a
polypeptide with at least one amino acid substitution relative to the polypeptide encoded by the first synthetic nucleic acid sequence.

59. The method of claim 55 wherein the synthetic sequences encode a luciferase.

60-64. (canceled)

65. The method of claim 48 or 49 further comprising altering the further synthetic nucleic acid molecule to encode a polypeptide having at least one amino acid substitution relative to the polypeptide encoded by the parent nucleic acid sequence.

66. The method of claim 48 or 49 wherein the altering of transcription regulatory sequences does not introduce amino acid substitutions to the polypeptide encoded by the synthetic nucleic acid molecule.

* * * * *