It has been discovered that certain natural mRNAs serve as metabolite-sensitive genetic switches wherein the RNA directly binds a small organic molecule. This binding process changes the conformation of the mRNA, which causes a change in gene expression by a variety of different mechanisms. Modified versions of these natural "riboswitches" (created by using various nucleic acid engineering strategies) can be employed as designer genetic switches that are controlled by specific effector compounds. Such effector compounds that activate a riboswitch are referred to herein as trigger molecules. The natural switches are targets for antibiotics and other small molecule therapies. In addition, the architecture of riboswitches allows actual pieces of the natural switches to be used to construct new non-immunogenic genetic control elements, for example the aptamer (molecular recognition) domain can be swapped with other non-natural aptamers (or otherwise modified) such that the new recognition domain causes genetic modulation with user-defined effector compounds. The changed switches become part of a therapy regimen—turning on, or off, or regulating protein synthesis. Newly constructed genetic regulation networks can be applied in such areas as living biosensors, metabolic engineering of organisms, and in advanced forms of gene therapy treatments.
FIG. 5E

FIG. 5F
0 CONSTANT SCISSION
0 INCREASED SCISSION
0 REDUCED SCISSION

thi box

FIG. 6B
○ CONSTANT SCISSON
○ INCREASED SCISSON
□ REDUCED SCISSION

FIG. 6D
FIG. 8A
FIG. 8B
FIG. 9A
<table>
<thead>
<tr>
<th>mutant</th>
<th>TPP binding</th>
<th>SD status (+ TPP)</th>
<th>genetic modulation (−/+ TPP ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>yes</td>
<td>closes</td>
<td>18</td>
</tr>
<tr>
<td>M1</td>
<td>no</td>
<td>unchanged</td>
<td>1.1</td>
</tr>
<tr>
<td>M2</td>
<td>yes</td>
<td>closes</td>
<td>16</td>
</tr>
<tr>
<td>M3</td>
<td>no</td>
<td>unchanged</td>
<td>1.1</td>
</tr>
<tr>
<td>M4</td>
<td>yes</td>
<td>closes</td>
<td>4.8</td>
</tr>
<tr>
<td>M5</td>
<td>no</td>
<td>unchanged</td>
<td>2.1</td>
</tr>
<tr>
<td>M6</td>
<td>yes</td>
<td>n.d.</td>
<td>10</td>
</tr>
<tr>
<td>M7</td>
<td>yes</td>
<td>n.d.</td>
<td>4.1</td>
</tr>
<tr>
<td>M8</td>
<td>yes</td>
<td>n.d.</td>
<td>1.6</td>
</tr>
<tr>
<td>M9</td>
<td>yes</td>
<td>n.d.</td>
<td>2.4</td>
</tr>
</tbody>
</table>

FIG.9C
SAM Aptamer

Riboswitch-Ribozyme Chimera

FIG. 10
FIG. 11D

FIG. 11E
purine aptamer consensus (adenine specific)

FIG. 11F
FIG. 12A
FIG. 12C
FIG. 16B

fraction bound

log c (TPP, M)

apparent K_D

~50 nM
Translation Control

TPP-induced alteration of 5' UTR secondary structure sequesters the Shine-Dalgarno (SD) sequence and prevents ribosome binding.

Transcriptional Control

TPP-induced alteration of 5' UTR secondary structure allows formation of a terminator hairpin, thus inducing transcription termination.

Splicing Control?

TPP binding domain is located in the first intron of a putative thiamine biosynthetic gene. TPP binding might control splicing of pre-mRNAs.

Processing/Stability Control?

TPP binding domain is located immediately adjacent to the polyA tail of a putative thiamine biosynthetic gene. TPP binding might control mRNA processing or stability.

FIG. 17

E. coli

5'UTR

5'UTR

3'UTR

3'UTR

thiM (THZ Kinase)

AGGAGC

AGGAGC

AUG

AUG

5' splice site

5' splice site

F. oxysporum

5' splice site

3' splice site

B. subtilis

5'UTR

thiamine biosynthetic gene

3'UTR

3'UTR

polYA tail
FIG. 20C

1. 6N HCl, 110°C 24hr
2. Neutralize

amino acids

active

inactive

ala

al

lys

ala


al


dipeptides

inactive

inactive

lys-ala

ala-ala
FIG. 20D

FIG. 20E
<table>
<thead>
<tr>
<th>Miller Units</th>
<th>Genotype</th>
<th>WT</th>
<th>G39C</th>
<th>G40C</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
<td>11</td>
<td>216</td>
<td>224</td>
</tr>
<tr>
<td>=</td>
<td></td>
<td>82</td>
<td>32</td>
<td>233</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>87</td>
<td>204</td>
<td>191</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>17</td>
<td>235</td>
<td>152</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>98</td>
<td>235</td>
<td>152</td>
</tr>
</tbody>
</table>

**FIG. 22C**

**FIG. 22B**
possible pseudoknot

G Box Consensus

FIG. 24A
FIG. 25A
inactive is adenine (A) >300,000 nM guanine hypoxanthine xanthine (G) (H) (X) KD 35 nM

FIG. 26A

FIG. 26C

= H bond acceptor

= H bond donor
a = $^3$H-guanine

$\frac{\text{fraction cpm in chamber \textit{b}}}{\text{distribution shift}} \sim 1.0$

$\sim 0.5$

$\sim 0.5$

$b = \text{no RNA (or)}$

$b = \text{WT RNA (or)}$

$b = \text{defective RNA mutant (or)}$

$b = \text{functional RNA mutant (or)}$

$b = \text{WT RNA + competitor}$

$b = \text{WT RNA + non-competitor}$

FIG. 27A

unlabeled analog concentration

0.5 $\mu$M

10 $\mu$M

frac. cpm counted in chamber \textit{b}

analog: - G H X A Ino Gua H X

RNA: none 93 xpt

FIG. 27B
FIG. 31A
124 yitJ aptamer domain

FIG. 32A
FIG. 33B

<table>
<thead>
<tr>
<th></th>
<th>Ma</th>
<th>Mab</th>
<th>Mc</th>
<th>Mabc</th>
<th>WT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>% terminated</td>
<td>22</td>
<td>25</td>
<td>64</td>
<td>74</td>
<td>83</td>
</tr>
</tbody>
</table>

FL
T
SAM
FIG. 33C
FROM FIG. 55B

FIG. 33D
FIG. 37A
Figure 37B: Graph showing the relationship between the log concentration of adenine (M) and the fraction cleaved (normalized). The graph includes data points for different sites (1-4) with a dissociation constant, $K_D$, of approximately 300 nM.
FIG. 39A

93 xpt (wild type)

fraction cleaved (normalized)

log c (ligand, M)

FIG. 39B

93 xpt (C to U)

fraction cleaved (normalized)

log c (ligand, M)
FIG. 40A
### Table S1. S Box Sequence Alignment

<table>
<thead>
<tr>
<th>ID</th>
<th>Position</th>
<th>Genbank Acc.</th>
<th>Organism</th>
<th>Remark</th>
<th>Start Operon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bs01</td>
<td>1160274</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>92</td>
<td>metF</td>
</tr>
<tr>
<td>Bs02</td>
<td>1257777</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>70</td>
<td>metB-metC</td>
</tr>
<tr>
<td>Bs03</td>
<td>1355353</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>130</td>
<td>metE</td>
</tr>
<tr>
<td>Bs04</td>
<td>1424147</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>(*)</td>
<td>ykrT-GCN3</td>
</tr>
<tr>
<td>Bs05</td>
<td>1426344</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>60</td>
<td>rbcl-ykrX-araD-ykrZ</td>
</tr>
<tr>
<td>Bs06</td>
<td>1629516</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>164</td>
<td>cysH-pitA-MET3-cysC</td>
</tr>
<tr>
<td>Bs07</td>
<td>2024504</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>86</td>
<td>TdhA-xy1B</td>
</tr>
<tr>
<td>Bs08</td>
<td>3128412</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>170</td>
<td>metK</td>
</tr>
<tr>
<td>Bs09</td>
<td>3360560</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>108</td>
<td>abc-2011-nlpA</td>
</tr>
<tr>
<td>Bs10</td>
<td>3995669</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>85</td>
<td>metE</td>
</tr>
<tr>
<td>Bs11</td>
<td>3997959</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
<td>80</td>
<td>metE</td>
</tr>
<tr>
<td>Bn01</td>
<td>910190</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
<td>141</td>
<td>???</td>
</tr>
<tr>
<td>Bn02</td>
<td>1348818</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
<td>99</td>
<td>thrA</td>
</tr>
<tr>
<td>Bn03</td>
<td>1699959</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
<td>175</td>
<td>metB-metC-metF-metH</td>
</tr>
</tbody>
</table>

**FIG.41A**
<table>
<thead>
<tr>
<th>Biofilm</th>
<th>Accession</th>
<th>Strain Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bh04</td>
<td>3427466</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>Bh05</td>
<td>3591166</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>O101</td>
<td>727028</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O102</td>
<td>1098097</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O103</td>
<td>131943</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O104</td>
<td>2134364</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O105</td>
<td>2365511</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O106</td>
<td>2437305</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O107</td>
<td>2708643</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O108</td>
<td>2856863</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O109</td>
<td>3162075</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O110</td>
<td>3200636</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O111</td>
<td>3200766</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O112</td>
<td>3294474</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>O113</td>
<td>3466518</td>
<td>NC_004193.1</td>
<td>Oceanobacillus iheyensis</td>
</tr>
<tr>
<td>Ca01</td>
<td>453565</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Ca02</td>
<td>671354</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Ca03</td>
<td>1073886</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Ca04</td>
<td>1131539</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
</tbody>
</table>

FIG. 41B
<table>
<thead>
<tr>
<th>Strain</th>
<th>Accession</th>
<th>Species/Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca05</td>
<td>1976373</td>
<td>NC_003030.1 Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Ca06</td>
<td>2914839</td>
<td>NC_003030.1 Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Ca07</td>
<td>2991405</td>
<td>NC_003030.1 Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Cp01</td>
<td>2500081</td>
<td>NC_003366.1 Clostridium perfringens</td>
</tr>
<tr>
<td>Cp02</td>
<td>2665229</td>
<td>NC_003366.1 Clostridium perfringens</td>
</tr>
<tr>
<td>Lm01</td>
<td>137135</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm02</td>
<td>309383</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm03</td>
<td>637924</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm04</td>
<td>882772</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm05</td>
<td>1716649</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm06</td>
<td>1739595</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Lm07</td>
<td>2491174</td>
<td>NC_003210.1 Listeria monocytogenes</td>
</tr>
<tr>
<td>Li01</td>
<td>172401</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li02</td>
<td>327333</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li03</td>
<td>636911</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li04</td>
<td>871751</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li05</td>
<td>1772459</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li06</td>
<td>1790189</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Li07</td>
<td>2538251</td>
<td>NC_003212.1 Listeria innocua (*)</td>
</tr>
<tr>
<td>Sa01</td>
<td>15958</td>
<td>NC_002745.1 Staphylococcus aureus</td>
</tr>
</tbody>
</table>

**FIG.41C**
<table>
<thead>
<tr>
<th></th>
<th>Accession</th>
<th>Organism</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sa02</td>
<td>875385</td>
<td><em>Staphylococcus aureus</em></td>
<td>91</td>
<td>abc-2011-nlpA</td>
</tr>
<tr>
<td>Sa03</td>
<td>1844603</td>
<td><em>Staphylococcus aureus</em></td>
<td>108</td>
<td>metK</td>
</tr>
<tr>
<td>Sa04</td>
<td>2381620</td>
<td><em>Staphylococcus aureus</em></td>
<td>13</td>
<td>nhaC</td>
</tr>
<tr>
<td>Sc01</td>
<td>4708438</td>
<td><em>Streptomyces coelicolor</em></td>
<td>26</td>
<td>rC-moadth</td>
</tr>
</tbody>
</table>

**FIG. 41D**
<table>
<thead>
<tr>
<th>ID</th>
<th>Position</th>
<th>Genbank Acc.</th>
<th>Organism</th>
<th>Remark</th>
<th>Start Operon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ct01</td>
<td>+</td>
<td>606192</td>
<td>AE_006470</td>
<td>Chlorobium tepidum</td>
<td>107 CAC5-MET2</td>
</tr>
<tr>
<td>Tt01</td>
<td>+</td>
<td>500245</td>
<td>NC_003869.1</td>
<td>Thermoanaerobacter tengcongensis</td>
<td>55 metK</td>
</tr>
<tr>
<td>Tt02</td>
<td>-</td>
<td>1750367</td>
<td>NC_003869.1</td>
<td>Thermoanaerobacter tengcongensis</td>
<td>66 metF-metH-ehsC</td>
</tr>
<tr>
<td>Tt03</td>
<td>-</td>
<td>2076680</td>
<td>NC_003969.1</td>
<td>Thermoanaerobacter tengcongensis</td>
<td>78 thrA-CAC5</td>
</tr>
<tr>
<td>Fn01</td>
<td>-</td>
<td>987483</td>
<td>NC_003454.1</td>
<td>Fusobacterium nucleatum</td>
<td>84 metK</td>
</tr>
<tr>
<td>Fn02</td>
<td>-</td>
<td>1317650</td>
<td>NC_003454.1</td>
<td>Fusobacterium nucleatum</td>
<td>91 abc-2011-nipA</td>
</tr>
<tr>
<td>Dr01</td>
<td>+</td>
<td>1363063</td>
<td>NC_001263.1</td>
<td>Deinococcus radiodurans, chr 1</td>
<td>156 abc-2011-nipA-nipA</td>
</tr>
<tr>
<td>Dr02</td>
<td>+</td>
<td>980704</td>
<td>NC_001263.1</td>
<td>Deinococcus radiodurans, chr 1</td>
<td>41 metH-????-metF</td>
</tr>
<tr>
<td>Xa01</td>
<td>-</td>
<td>3379769</td>
<td>NC_003919.1</td>
<td>Xanthomonas axonopodis</td>
<td>74 MET2-metC-thrA</td>
</tr>
<tr>
<td>Xc01</td>
<td>-</td>
<td>3379769</td>
<td>NC_003902.1</td>
<td>Xanthomonas campestris</td>
<td>73 MET2-metC-thrA</td>
</tr>
<tr>
<td>Se01</td>
<td>+</td>
<td>574</td>
<td>AF_269983.1</td>
<td>Staphylococcus epidermidis genomic clone</td>
<td></td>
</tr>
<tr>
<td>Se02</td>
<td>-</td>
<td>142</td>
<td>AF_270301.1</td>
<td>Staphylococcus epidermidis genomic clone</td>
<td></td>
</tr>
<tr>
<td>Gs01</td>
<td>+</td>
<td>342043</td>
<td>contig:2947</td>
<td>Geobacter sulfurreducens</td>
<td></td>
</tr>
<tr>
<td>Gs02</td>
<td>+</td>
<td>2470946</td>
<td>contig:2947</td>
<td>Geobacter sulfurreducens</td>
<td></td>
</tr>
<tr>
<td>Ba01</td>
<td>-</td>
<td>177272</td>
<td>contig:6615</td>
<td>Bacillus anthracis</td>
<td></td>
</tr>
<tr>
<td>Ba02</td>
<td>+</td>
<td>185586</td>
<td>contig:6615</td>
<td>Bacillus anthracis</td>
<td></td>
</tr>
<tr>
<td>Ba03</td>
<td>-</td>
<td>197185</td>
<td>contig:6615</td>
<td>Bacillus anthracis</td>
<td></td>
</tr>
<tr>
<td>Ba04</td>
<td>+</td>
<td>320607</td>
<td>contig:6615</td>
<td>Bacillus anthracis</td>
<td></td>
</tr>
</tbody>
</table>

**FIG.41E**
Ba05  -  371127  contig:6615  Bacillus anthracis
Ba06  +  1362659  contig:6615  Bacillus anthracis
Ba07  +  1375353  contig:6615  Bacillus anthracis
Ba08  +  2459362  contig:6615  Bacillus anthracis
Ba09  -  2953226  contig:6615  Bacillus anthracis
Ba10  -  3091676  contig:6615  Bacillus anthracis
Ba11  -  3890736  contig:6615  Bacillus anthracis
Ba12  +  3892933  contig:6615  Bacillus anthracis
Ba13  +  4074285  contig:6615  Bacillus anthracis
Ba14  -  4074078  contig:6615  Bacillus anthracis
Ba15  -  4553682  contig:6615  Bacillus anthracis
Ba16  -  4739975  contig:6615  Bacillus anthracis  
Ba17  -  5140322  contig:6615  Bacillus anthracis
Bc01  +  26115  contig:1617  Bacillus anthracis  
Bc02  +  748841  contig:1617  Bacillus cereus
Bc03  +  11803078  contig:1617  Bacillus cereus  
Bc04  +  1195849  contig:1617  Bacillus cereus  
Bc05  +  2200884  contig:1617  Bacillus cereus  
Bc06  -  2684484  contig:1617  Bacillus cereus  
Bc07  -  2963724  contig:1617  Bacillus cereus

FIG.41F
FIG. 41J
FIG.41M
## B. Cobalamin Riboswitch Alignment

<table>
<thead>
<tr>
<th>ID</th>
<th>Start</th>
<th>End</th>
<th>Accession</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atu01</td>
<td>70441</td>
<td>70625</td>
<td>NC_003304.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Atu02</td>
<td>441331</td>
<td>441136</td>
<td>NC_003305.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Atu03</td>
<td>877645</td>
<td>877833</td>
<td>NC_003304.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Atu04</td>
<td>921717</td>
<td>921806</td>
<td>NC_003305.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Atu05</td>
<td>1640563</td>
<td>1640420</td>
<td>NC_003304.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Atu06</td>
<td>2810076</td>
<td>2809899</td>
<td>NC_003304.1</td>
<td>Agrobacterium tumefaciens</td>
</tr>
<tr>
<td>Bhal01</td>
<td>466904</td>
<td>466746</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>Bhal02</td>
<td>528894</td>
<td>529051</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>Bhal03</td>
<td>870599</td>
<td>870748</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>Bhal04</td>
<td>1661078</td>
<td>1661219</td>
<td>NC_002570.1</td>
<td>Bacillus halodurans</td>
</tr>
<tr>
<td>Bsu01</td>
<td>3403719</td>
<td>3403620</td>
<td>NC_000964.1</td>
<td>Bacillus subtilis</td>
</tr>
<tr>
<td>Bja01</td>
<td>2232013</td>
<td>2232975</td>
<td>NC_004463.1</td>
<td>Bradyrhizobium japonicum</td>
</tr>
<tr>
<td>Bja02</td>
<td>3617311</td>
<td>3617490</td>
<td>NC_004463.1</td>
<td>Bradyrhizobium japonicum</td>
</tr>
<tr>
<td>Bja03</td>
<td>3630677</td>
<td>3630857</td>
<td>NC_004463.1</td>
<td>Bradyrhizobium japonicum</td>
</tr>
<tr>
<td>Bja04</td>
<td>3634122</td>
<td>3634284</td>
<td>NC_004463.1</td>
<td>Bradyrhizobium japonicum</td>
</tr>
<tr>
<td>Bja05</td>
<td>5279669</td>
<td>5279495</td>
<td>NC_004463.1</td>
<td>Bradyrhizobium japonicum</td>
</tr>
<tr>
<td>Bme01</td>
<td>679030</td>
<td>679218</td>
<td>NC_003317.1</td>
<td>Brucella melitensis</td>
</tr>
<tr>
<td>Bme02</td>
<td>717388</td>
<td>717585</td>
<td>NC_003317.1</td>
<td>Brucella melitensis</td>
</tr>
</tbody>
</table>

**FIG.41X**
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bme03</td>
<td>+</td>
<td>559758</td>
<td>559950</td>
<td>NC_003318.1</td>
<td>Brucella melitensis chromosome II</td>
</tr>
<tr>
<td>Bme04</td>
<td>-</td>
<td>973106</td>
<td>972933</td>
<td>NC_003318.1</td>
<td>Brucella melitensis chromosome II</td>
</tr>
<tr>
<td>Ccr01</td>
<td>+</td>
<td>502968</td>
<td>503156</td>
<td>NC_002696.2</td>
<td>Caulobacter crescentus</td>
</tr>
<tr>
<td>Ccr02</td>
<td>+</td>
<td>1925017</td>
<td>1925166</td>
<td>NC_002696.2</td>
<td>Caulobacter crescentus</td>
</tr>
<tr>
<td>Cte01</td>
<td>-</td>
<td>409802</td>
<td>409630</td>
<td>NC_002932.3</td>
<td>Chlorobium tepidum</td>
</tr>
<tr>
<td>Cte02</td>
<td>+</td>
<td>422045</td>
<td>422244</td>
<td>NC_002932.3</td>
<td>Chlorobium tepidum</td>
</tr>
<tr>
<td>Cte03</td>
<td>+</td>
<td>443769</td>
<td>443951</td>
<td>NC_002932.3</td>
<td>Chlorobium tepidum</td>
</tr>
<tr>
<td>Cte04</td>
<td>+</td>
<td>584183</td>
<td>584411</td>
<td>NC_002932.3</td>
<td>Chlorobium tepidum</td>
</tr>
<tr>
<td>Cte05</td>
<td>+</td>
<td>882576</td>
<td>882770</td>
<td>NC_002932.3</td>
<td>Chlorobium tepidum</td>
</tr>
<tr>
<td>Cac01</td>
<td>+</td>
<td>1509969</td>
<td>1510116</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Cac02</td>
<td>+</td>
<td>2557903</td>
<td>2558041</td>
<td>NC_003030.1</td>
<td>Clostridium acetobutylicum</td>
</tr>
<tr>
<td>Cpe01</td>
<td>+</td>
<td>248269</td>
<td>248429</td>
<td>NC_003366.1</td>
<td>Clostridium perfringens</td>
</tr>
<tr>
<td>Cpe02</td>
<td>+</td>
<td>1241749</td>
<td>1241887</td>
<td>NC_003366.1</td>
<td>Clostridium perfringens</td>
</tr>
<tr>
<td>Cpe03</td>
<td>-</td>
<td>1431291</td>
<td>1431152</td>
<td>NC_003366.1</td>
<td>Clostridium perfringens</td>
</tr>
<tr>
<td>Cpe04</td>
<td>-</td>
<td>1549481</td>
<td>1549348</td>
<td>NC_003366.1</td>
<td>Clostridium perfringens</td>
</tr>
<tr>
<td>Eco01</td>
<td>+</td>
<td>4160983</td>
<td>4161133</td>
<td>NC_000913.1</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>Fnu01</td>
<td>+</td>
<td>934517</td>
<td>934658</td>
<td>NC_003454.1</td>
<td>Fusobacterium nucleatum</td>
</tr>
<tr>
<td>Lig01</td>
<td>+</td>
<td>1347854</td>
<td>1347994</td>
<td>NC_004342.1</td>
<td>Leptospira interrogans chromosome I</td>
</tr>
<tr>
<td>Lmo01</td>
<td>+</td>
<td>1179829</td>
<td>1179979</td>
<td>NC_003210.1</td>
<td>Listeria monocytogenes</td>
</tr>
<tr>
<td>Mlo01</td>
<td>-</td>
<td>1101076</td>
<td>1100918</td>
<td>NC_002678.1</td>
<td>Mesorhizobium loti</td>
</tr>
</tbody>
</table>

**FIG.41Y**
<table>
<thead>
<tr>
<th>Mi02</th>
<th>+</th>
<th>1149143 1149308</th>
<th>NC_002678.1</th>
<th>Mesorhizobium loti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mi003</td>
<td>-</td>
<td>4044577 4044416</td>
<td>NC_002678.1</td>
<td>Mesorhizobium loti</td>
</tr>
<tr>
<td>Mi004</td>
<td>-</td>
<td>4957334 4957164</td>
<td>NC_002678.1</td>
<td>Mesorhizobium loti</td>
</tr>
<tr>
<td>Mi005</td>
<td>-</td>
<td>6170855 6170715</td>
<td>NC_002678.1</td>
<td>Mesorhizobium loti</td>
</tr>
<tr>
<td>Mi006</td>
<td>+</td>
<td>6749148 6749315</td>
<td>NC_002678.1</td>
<td>Mesorhizobium loti</td>
</tr>
<tr>
<td>Mle01</td>
<td>-</td>
<td>1130394 1130222</td>
<td>NC_002677.1</td>
<td>Mycobacterium leprae</td>
</tr>
<tr>
<td>Mtu01</td>
<td>-</td>
<td>309822 309703</td>
<td>NC_000962.1</td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>Mtu02</td>
<td>-</td>
<td>1261701 1261497</td>
<td>NC_000962.1</td>
<td>Mycobacterium tuberculosis</td>
</tr>
<tr>
<td>Pae01</td>
<td>+</td>
<td>1381520 1381688</td>
<td>NC_002516.1</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>Pae02</td>
<td>-</td>
<td>3261415 3261204</td>
<td>NC_002516.1</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>Pae03</td>
<td>+</td>
<td>3265553 3265728</td>
<td>NC_002516.1</td>
<td>Pseudomonas aeruginosa</td>
</tr>
</tbody>
</table>

**FIG.41Z**
<table>
<thead>
<tr>
<th>ID</th>
<th>Start</th>
<th>End</th>
<th>Accession</th>
<th>Organism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pae04</td>
<td>3305780</td>
<td>3305629</td>
<td>NC_002516.1</td>
<td>Pseudomonas aeruginosa</td>
</tr>
<tr>
<td>Ppu01</td>
<td>2756203</td>
<td>2756045</td>
<td>NC_002947.3</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Ppu02</td>
<td>2756853</td>
<td>2756785</td>
<td>NC_002947.3</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Ppu03</td>
<td>3815563</td>
<td>3857746</td>
<td>NC_002947.3</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Ppu04</td>
<td>3981958</td>
<td>3981816</td>
<td>NC_002947.3</td>
<td>Pseudomonas putida</td>
</tr>
<tr>
<td>Rso01</td>
<td>2609233</td>
<td>2609017</td>
<td>NC_003295.1</td>
<td>Ralstonia solanacearum</td>
</tr>
<tr>
<td>Sme01</td>
<td>954780</td>
<td>954943</td>
<td>NC_003047.1</td>
<td>Sinorhizobium meliloti</td>
</tr>
<tr>
<td>Sme02</td>
<td>1999747</td>
<td>1999674</td>
<td>NC_003047.1</td>
<td>Sinorhizobium meliloti</td>
</tr>
<tr>
<td>Sme03</td>
<td>2122891</td>
<td>2122516</td>
<td>NC_003047.1</td>
<td>Sinorhizobium meliloti</td>
</tr>
<tr>
<td>Sme04</td>
<td>66265</td>
<td>66456</td>
<td>NC_003078.1</td>
<td>Sinorhizobium meliloti plasmid pSymB</td>
</tr>
<tr>
<td>Sme05</td>
<td>580403</td>
<td>580578</td>
<td>NC_003078.1</td>
<td>Sinorhizobium meliloti plasmid pSymB</td>
</tr>
<tr>
<td>Sco01</td>
<td>1037869</td>
<td>1038053</td>
<td>NC_003888.1</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Sco02</td>
<td>1048899</td>
<td>1046106</td>
<td>NC_003888.1</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Sco03</td>
<td>105420</td>
<td>1051563</td>
<td>NC_003888.1</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Sco04</td>
<td>5688395</td>
<td>5688291</td>
<td>NC_003888.1</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Sco05</td>
<td>6532337</td>
<td>6532191</td>
<td>NC_003888.1</td>
<td>Streptomyces coelicolor</td>
</tr>
<tr>
<td>Sf101</td>
<td>4183416</td>
<td>4183566</td>
<td>NC_004337.1</td>
<td>Shigella flexneri (*)</td>
</tr>
<tr>
<td>Ssc01</td>
<td>826836</td>
<td>827026</td>
<td>NC_004347.1</td>
<td>Shewanella oneidensis</td>
</tr>
</tbody>
</table>

FIG. 41AA
<table>
<thead>
<tr>
<th>Code</th>
<th>Start</th>
<th>End</th>
<th>Accession</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Son02</td>
<td>1071692</td>
<td>1071874</td>
<td>NC_004347.1</td>
<td>Shewanella oneidensis</td>
</tr>
<tr>
<td>Sti01</td>
<td>2114053</td>
<td>2113918</td>
<td>NC_003197.1</td>
<td>Salmonella typhimurium</td>
</tr>
<tr>
<td>Sti02</td>
<td>4347871</td>
<td>4348024</td>
<td>NC_003197.1</td>
<td>Salmonella typhimurium</td>
</tr>
<tr>
<td>Tma01</td>
<td>84288</td>
<td>84144</td>
<td>NC_000853.1</td>
<td>Thermotoga maritima</td>
</tr>
<tr>
<td>Tte01</td>
<td>395153</td>
<td>395353</td>
<td>NC_003869.1</td>
<td>Thermoanaerobacter tengcongensis</td>
</tr>
<tr>
<td>Tte02</td>
<td>396075</td>
<td>396275</td>
<td>NC_003869.1</td>
<td>Thermoanaerobacter tengcongensis (*)</td>
</tr>
<tr>
<td>Vch01</td>
<td>145142</td>
<td>145306</td>
<td>NC_002505.1</td>
<td>Vibrio cholerae chromosome I</td>
</tr>
<tr>
<td>Vvu01</td>
<td>1165724</td>
<td>1165882</td>
<td>NC_004459.1</td>
<td>Vibrio vulnificus chromosome I</td>
</tr>
<tr>
<td>Xac01</td>
<td>3631166</td>
<td>3630987</td>
<td>NC_003902.1</td>
<td>Xanthomonas campestris</td>
</tr>
<tr>
<td>Xax01</td>
<td>3758428</td>
<td>3758245</td>
<td>NC_003919.1</td>
<td>Xanthomonas citri</td>
</tr>
<tr>
<td>Ype01</td>
<td>4393235</td>
<td>4393008</td>
<td>NC_003143.1</td>
<td>Yersinia pestis</td>
</tr>
<tr>
<td>Aca01</td>
<td>340</td>
<td>170</td>
<td>M34485.1</td>
<td>Acinetobacter calcoaceticus</td>
</tr>
<tr>
<td>Avi01</td>
<td>388</td>
<td>214</td>
<td>U45329.1</td>
<td>Agrobacterium vitis</td>
</tr>
<tr>
<td>Bfr01</td>
<td>580</td>
<td>762</td>
<td>AY043208.1</td>
<td>Bacteroides fragilis</td>
</tr>
<tr>
<td>Bmg01</td>
<td>1211</td>
<td>1350</td>
<td>AJ000758.1</td>
<td>Bacillus megaterium</td>
</tr>
<tr>
<td>Lma01</td>
<td>76392</td>
<td>76234</td>
<td>AL499620.1</td>
<td>Leishmania major</td>
</tr>
<tr>
<td>Pfr01</td>
<td>543</td>
<td>373</td>
<td>AY033236.1</td>
<td>Propionibacterium freudenreichii</td>
</tr>
<tr>
<td>Rca01</td>
<td>105327</td>
<td>105521</td>
<td>AF010496.1</td>
<td>Rhodobacter capsulatus</td>
</tr>
<tr>
<td>Rca02</td>
<td>116991</td>
<td>117174</td>
<td>AF010496.1</td>
<td>Rhodobacter capsulatus</td>
</tr>
<tr>
<td>Rca03</td>
<td>39849</td>
<td>39672</td>
<td>AF010496.1</td>
<td>Rhodobacter capsulatus</td>
</tr>
</tbody>
</table>

**FIG. 41AB**
Accession numbers are for Genbank sequence entries. Start and end positions are the 5' and 3' nucleotides of the first interior UG base pair of stem PI (orange). Secondary structure (SS) and sequence consensus (Cons) lines are shown above the alignment. In the consensus sequence, uppercase and lowercase letters represent >=90% and >=80% conservation at a position, respectively. The degenerate bases R (A,G) and Y (C,U) appear only when a single base is not 80% conserved. Sequences marked with an asterisk (*) were excluded when determining the sequence consensus because they have >90% identity to another sequence in the alignment.

FIG.41AC
FIG. 41AH
FIG. 41 AT
FIG. 41AU
NC_002745.1/430771-430861  GTAAAAATAGCATAAAAATTCCGataATCTAAAGAataatgctTTTAGAAgtttctctcCTGACTGccTggGaaCCACATG, actaTGAGAAACGATCCATTTAG
Staphylococcus aureus subsp. aureus

NC_004461.1/2432364-2432294  CATAAAAATATTTATAGCAGCATAaATGTAGAGAataatgctTTTAGAAgtttctccCTGTCGGccAttGaaCCACAG, actaTGAGAAACGATCCATTTAG
Staphylococcus epidermidis

NC_004116.1/1093950-1093860  CAATTAAATATATAGTTTACTTTATTtaT, GCTGAGGat, tgg, CTGACGatCtctacAGAGAC, ccGl, aG, TGCTA, acAATAAGTTGCAATGAAATGCT
Streptococcus agalactiae

NC_002737.1/930757-930842  TGAATCTCAAATAGCTACTTATTtaT, GCTGAGGat, tgg, CCAAGGgtCtctacAGAGAC, cc.ttaa, TGCTA, acAATAAGTTGCAATGAAATGCT
Streptococcus pyogenes

NC_003028.1/1754791-1754878  AAAAAATATATCGTTTACTTTGGTTTatT, GTGGTaat, tgg, CACCAGatTtctacAGGTG, cc.Ggaa, CAGCT, AacAATAAGTTGCAATGAAATGCT
Streptococcus pneumoniae

NC_003969.1/586372 586463  AAAATTTAAATAAGCAGCATAAATCCGAGAtatgct, CTGAGGatCtctacAGAGAC, ccGttaaTiGTCGG, actaTGAGAAACGATCCATTTAG
Thermoanaerobacter tengcongensis

Cconsensus

............<--------<----------<----------<----------<----------<----------<----------<----------<----------<----------<----------<----------

**********Y*TAVTA**********AT**GG**********GT*YCTA**********CC***AA**********WAVR*R****

FIG.41AV
E. Lysine riboswitch comparison

Command-1: Plain Text

1. Command-2: Base paired stem 1
2. Command-3: Base paired stem 2
   a. Command-4: Base paired stem 3
3. Command-5: Base paired stem 4
4. Command-6: Base paired stem 5
   b. Command-7: Base paired stem 6
5. Command-8: Terminator poly-U
6. Command-9: Downstream AT stem paired to stem 1
7. Command-0: Optional base paired stem 2

CUAG is 90% sequence similarity

2. CUAG is the Anti-Terminator

CUAG is the Terminator stem

1. Bha LysC  AGUGAUGUlagaggU-gcGAAAACC-aAG-aguaC-ACAAGUGAAGAAAUG---AGAAU---CGUUGAC---GACUGUUGGAAagg--
GGAUUGCagcggaggUGAGAGAGGGG--CUCAUAGGG--AUUUUGGCUUlagAGCUAAGG---gaauUA-AGCAUAGGACgucaACACAUAG---GCCCAG-----
CUAGCUUGUlagggagACguCAUCACGU

2. Bha dapA  AGUGAAGUGUlagaggU-gcGAAAACC-aAG-aguaC-CAAGAUGAAGAAAUG---GAUGA---GUAAGA---UGCUUGAGAAGUUUUGUGG--GAAagg--
GGAUUGCagcggaggUGAGAGAGGGG--CUCAUAGGG--AUUUUGGCUUlagAGCUAAGG---gaauUA-AGCAUAGGACgucaACACAUAG---GCAU-----
UGCUUUGUlagggagGCUAUCACAGC

AAAAUGCagcggaggCCGAAGAGUG---AUCAAGC-AGUCAUGUUGCUUlagCGGUAGAUU---gaauUA-AAGUGUACACgucaACAG---AGAAU-----
GCGUGGagACguCAUCACGU

GGAGCGUGCagcggaggCAAUUGAACCC--CCAGC-GCAUUGUUGCCGlagCGGUAGAUU---gaauUA-AUGUGUACAGgucaACAGA-------UCAU-----
UGUGUGGagGCUAUGGUUGUG

5. Cac LysA  AGCGUGUGUlagaggU-gcGUGUAGUUGC-aAG-aguaA-GGCGUUGUUGAUGAUGAAGAAAUG---GAGAAUGUACACCGAGGA-------GUGCCG--GMAagg--
AAAAagg-GCGUUGUlagcggaggCAUGAAGAGUG---GAGCA---GUAUGUUGUUGCCGlagCGGUAGAUU---gaauUA-AUGUGUACAGgucaACAGA-------UCAU-----
UGUGUGGagGCUAUGGUUGUG

FIG. 41AX
FIG. 41AZ
RIBOSWITCHES, METHODS FOR THEIR USE, AND COMPOSITIONS FOR USE WITH RIBOSWITCHES

CROSS-REFERENCE TO RELATED APPLICATIONS

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH
[0002] This invention was made with government support under Grants NIH GM48858 and NIH GM559343 awarded by the National Institutes of Health, and Grant NSF EIA-0129939 awarded by the National Science Foundation. The government has certain rights in the invention.

REFERENCE TO SEQUENCE LISTING

FIELD OF THE INVENTION
[0004] The disclosed invention is generally in the field of gene expression and specifically in the area of regulation of gene expression.

BACKGROUND OF THE INVENTION
[0005] Precision genetic control is an essential feature of living systems, as cells must respond to a multitude of biochemical signals and environmental cues by varying genetic expression patterns. Most known mechanisms of genetic control involve the use of protein factors that sense chemical or physical stimuli and then modulate gene expression by selectively interacting with the relevant DNA or messenger RNA sequence. Proteins can adopt complex shapes and carry out a variety of functions that permit living systems to sense accurately their chemical and physical environments. Protein factors that respond to metabolites typically act by binding DNA to modulate transcription initiation (e.g. the lac repressor protein; Matthews, K. S., and Nichols, J. C., 1998, Prog. Nucleic Acids Res. Mol. Biol. 58, 127-164) or by binding RNA to control either transcription termination (e.g. the PyrR protein; Switzer, R. L., et al., 1999, Prog. Nucleic Acids Res. Mol. Biol. 62, 329-367) or translation (e.g. the TRAP protein; Babitzke, P., and Gollnick, P., 2001, J. Bacteriol. 183, 5795-5802). Protein factors responds to environmental stimuli by various mechanisms such as allosteric modulation or posttranslational modification, and are adept at exploiting these mechanisms to serve as highly responsive genetic switches (e.g. see Pianese, M., and Gujar, A. (2002). Genes and Signals. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

[0006] In addition to the widespread participation of protein factors in genetic control, it is also known that RNA can take an active role in genetic regulation. Recent studies have begun to reveal the substantial role that small non-coding RNAs play in selectively targeting mRNAs for destruction, which results in down-regulation of gene expression (e.g. see Hanlon, O. J. 2002, Nature 418, 244-251 and references therein). This process of RNA interference takes advantage of the ability of short RNAs to recognize the intended mRNA target selectively via Watson-Crick base complementation, after which the bound RNAs are destroyed by the action of proteins. RNAs are ideal agents for molecular recognition in this system because it is far easier to generate new target-specific RNA factors through evolutionary processes than it would be to generate protein factors with novel but highly specific RNA binding sites.


**BRIEF SUMMARY OF THE INVENTION**

[0010] It has been discovered that certain natural mRNAs serve as metabolite-sensitive genetic switches wherein the RNA directly binds a small organic molecule. This binding process changes the conformation of the mRNA, which causes a change in gene expression by a variety of different mechanisms. Modified versions of these natural "riboswitches" (created by using various nucleic acid engineering strategies) can be employed as designer genetic switches that are controlled by specific effector compounds. Such effector compounds that activate a riboswitch are referred to herein as trigger molecules. The natural switches are targets for antibiotics and other small molecule therapies. In addition, the architecture of riboswitches allows actual pieces of the natural switches to be used to construct new non-immunogenic genetic control elements, for example the aptamer (molecular recognition) domain can be swapped with other non-natural aptamers (or otherwise modified) such that the new recognition domain causes genetic modulation with user-defined effector compounds. The changed switches become part of a therapy regimen—turning on, or off, or regulating protein synthesis. Newly constructed genetic regulation networks can be applied in such areas as living biosensors, metabolic engineering of organisms, and in advanced forms of gene therapy treatments.

[0011] Disclosed are isolated and recombinant riboswitches, recombinant constructs containing such riboswitches, heterologous sequences operably linked to such riboswitches, and cells and transgenic organisms harboring such riboswitches, riboswitch recombinant constructs, and riboswitches operably linked to heterologous sequences. The heterologous sequences can be, for example, sequences encoding proteins or peptides of interest, including reporter proteins or peptides. Preferred riboswitches are, or are derived from, naturally occurring riboswitches.

[0012] Also disclosed are chimeric riboswitches containing heterologous aptamer domains and expression platform domains. That is, chimeric riboswitches are made up an aptamer domain from one source and an expression platform domain from another source. The heterologous sources can be from, for example, different specific riboswitches or different classes of riboswitches. The heterologous aptamers can also come from non-riboswitch aptamers. The heterologous expression platform domains can also come from non-riboswitch sources.

[0013] Also disclosed are compositions and methods for selecting and identifying compounds that can activate, deactivate or block a riboswitch. Activation of a riboswitch refers to the change in state of the riboswitch upon binding of a trigger molecule. A riboswitch can be activated by compounds other than the trigger molecule and in ways other than binding of a trigger molecule. The term trigger molecule is used herein to refer to molecules and compounds that can activate a riboswitch. This includes the natural or normal trigger molecule for the riboswitch and other compounds that can activate the riboswitch. Natural or normal trigger molecules are the trigger molecule for a given riboswitch in nature or, in the case of some non-natural riboswitches, the trigger molecule for which the riboswitch was designed or with which the riboswitch was selected (as in, for example, in vitro selection or in vitro evolution techniques). Non-natural trigger molecules can be referred to as non-natural trigger molecules.

[0014] Deactivation of a riboswitch refers to the change in state of the riboswitch when the trigger molecule is not bound. A riboswitch can be deactivated by binding of compounds other than the trigger molecule and in ways other than removal of the trigger molecule. Blocking of a riboswitch refers to a condition or state of the riboswitch where the presence of the trigger molecule does not activate the riboswitch.

[0015] Also disclosed are compounds, and compositions containing such compounds, that can activate, deactivate or block a riboswitch. Also disclosed are compositions and methods for activating, deactivating or blocking a riboswitch. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Compounds can be used to activate, deactivate or block a riboswitch. The trigger molecule for a riboswitch (as well as other activating compounds) can be used to activate a riboswitch. Compounds other than the trigger molecule generally can be used to deactivate or block a riboswitch. Riboswitches can also be deactivated by, for example, removing trigger molecules from the presence of the riboswitch. A riboswitch can be blocked by, for example, binding of an analog of the trigger molecule that does not activate the riboswitch.

[0016] Also disclosed are compositions and methods for altering expression of an RNA molecule, or of a gene encoding an RNA molecule, where the RNA molecule includes a riboswitch, by bringing a compound into contact with the RNA molecule. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Thus, subjecting an RNA molecule of interest that includes a riboswitch to conditions that activate, deactivate or block the riboswitch can be used to alter expression of the RNA. Expression can be altered as a result of, for example, termination of transcription or blocking of ribosome binding to the RNA. Binding of a trigger molecule can, depending on the nature of the riboswitch, reduce or prevent expression of the RNA molecule or promote or increase expression of the RNA molecule.

[0017] Also disclosed are compositions and methods for regulating expression of an RNA molecule, or of a gene encoding an RNA molecule, by operably linking a riboswitch to the RNA molecule. A riboswitch can be operably linked to an RNA molecule in any suitable manner, including, for example, by physically joining the riboswitch to the RNA molecule or by engineering nucleic acid encoding the RNA molecule to include and encode the riboswitch such that the RNA produced from the engineered nucleic acid has the riboswitch operably linked to the RNA molecule. Subjecting a riboswitch operably linked to an RNA molecule of interest
to conditions that activate, deactivate or block the riboswitch can be used to alter expression of the RNA.

[0018] Also disclosed are compositions and methods for regulating expression of a naturally occurring gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. If the gene is essential for survival of a cell or organism that harbors it, activating, deactivating or blocking the riboswitch can in death, stasis or debilitation of the cell or organism. For example, activating a naturally occurring riboswitch in a naturally occurring gene that is essential to survival of a microorganism can result in death of the microorganism (if activation of the riboswitch turns off or represses expression). This is one basis for the use of the disclosed compounds and methods for antimicrobial and antibiotic effects.

[0019] Also disclosed are compositions and methods for regulating expression of an isolated, engineered or recombinant gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. The gene or RNA can be engineered or can be recombinant in any manner. For example, the riboswitch and coding region of the RNA can be heterologous, the riboswitch can be recombinant or chimeric, or both. If the gene encodes a desired expression product, activating or deactivating the riboswitch can be used to induce expression of the gene and thus result in production of the expression product. If the gene encodes an inducer or repressor of gene expression or of another cellular process, activation, deactivation or blocking of the riboswitch can result in induction, repression, or de-repression of other, regulated genes or cellular processes. Many such secondary regulatory effects are known and can be adapted for use with riboswitches. An advantage of riboswitches as the primary control for such regulation is that riboswitch trigger molecules can be small, non-antibiotic molecules.

[0020] Also disclosed are compositions and methods for altering the regulation of a riboswitch by operably linking an aptamer domain to the expression platform domain of the riboswitch (which is a chimeric riboswitch). The aptamer domain can then mediate regulation of the riboswitch through the action of, for example, a trigger molecule for the aptamer domain. Aptamer domains can be operably linked to expression platform domains of riboswitches in any suitable manner, including, for example, by replacing the normal or natural aptamer domain of the riboswitch with the new aptamer domain. Generally, any compound or condition that can activate, deactivate or block the riboswitch from which the aptamer domain is derived can be used to activate, deactivate or block the chimeric riboswitch.

[0021] Also disclosed are compositions and methods for inactivating a riboswitch by covalently altering the riboswitch (by, for example, crosslinking parts of the riboswitch or coupling a compound to the riboswitch). Inactivation of a riboswitch in this manner can result from, for example, an alteration that prevents the trigger molecule for the riboswitch from binding, that prevents the change in state of the riboswitch upon binding of the trigger molecule, or that prevents the expression platform domain of the riboswitch from affecting expression upon binding of the trigger molecule.

[0022] Also disclosed are methods of identifying compounds that activate, deactivate or block a riboswitch. For examples, compounds that activate a riboswitch can be identified by bringing into contact a test compound and a riboswitch and assessing activation of the riboswitch. If the riboswitch is activated, the test compound is identified as a compound that activates the riboswitch. Activation of a riboswitch can be assessed in any suitable manner. For example, the riboswitch can be linked to a reporter RNA and expression, expression level, or change in expression level of the reporter RNA can be measured in the presence and absence of the test compound. As another example, the riboswitch can include a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch. As can be seen, assessment of activation of a riboswitch can be performed with the use of a control assay or measurement or without the use of a control assay or measurement. Methods for identifying compounds that deactivate a riboswitch can be performed in analogous ways.

[0023] Identification of compounds that block a riboswitch can be accomplished in any suitable manner. For example, an assay can be performed for assessing activation or deactivation of a riboswitch in the presence of a compound known to activate or deactivate the riboswitch and in the presence of a test compound. If activation or deactivation is not observed as would be observed in the absence of the test compound, then the test compound is identified as a compound that blocks activation or deactivation of the riboswitch.

[0024] Also disclosed are biosensor riboswitches. Biosensor riboswitches are engineered riboswitches that produce a detectable signal in the presence of their cognate trigger molecule. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Biosensor riboswitches can be designed for use in vivo or in vitro. For example, biosensor riboswitches operably linked to a reporter RNA that encodes a protein that serves as or is involved in producing a signal can be used in vivo by engineering a cell or organism to harbor a nucleic acid construct encoding the riboswitch/reporter RNA. An example of a biosensor riboswitch for use in vitro is a riboswitch that includes a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a biosensor riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch. Also disclosed are methods of detecting compounds using biosensor riboswitches. The method can include bringing into contact a test sample and a biosensor riboswitch and assessing the activation of the biosensor riboswitch. Activation of the biosensor riboswitch indicates the presence of the trigger molecule for the biosensor riboswitch in the test sample.

[0025] Also disclosed are compounds made by identifying a compound that activates, deactivates or blocks a riboswitch and manufacturing the identified compound. This can be accomplished by, for example, combining compound identification methods as disclosed elsewhere herein with methods for manufacturing the identified compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound.

[0026] Also disclosed are compounds made by checking activation, deactivation or blocking of a riboswitch by a compound and manufacturing the checked compound. This can be accomplished by, for example, combining compound activation, deactivation or blocking assessment methods as disclosed elsewhere herein with methods for manufacturing the compound.
checked compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound. Checking compounds for their ability to activate, deactivate or block a riboswitch refers to both identification of compounds previously unknown to activate, deactivate or block a riboswitch and to assessing the ability of a compound to activate, deactivate or block a riboswitch where the compound was already known to activate, deactivate or block the riboswitch.

[0027] Also disclosed are methods for selecting, designing or deriving new riboswitches and/or new aptamers that recognize new trigger molecules. Such methods can involve production of a set of aptamer variants in a riboswitch, assessing the activation of the variant riboswitches in the presence of a compound of interest, selecting variant riboswitches that were activated (or, for example, the riboswitches that were the most highly or the most selectively activated), and repeating these steps until a variant riboswitch of a desired activity, specificity, combination of activity and specificity, or other combination of properties results. Also disclosed are riboswitches and aptamer domains produced by these methods.

[0028] The disclosed riboswitches, including the derivatives and recombinant forms thereof, generally can be from any source, including naturally occurring riboswitches and riboswitches designed de novo. Any such riboswitches can be used in or with the disclosed methods. However, different types of riboswitches can be defined and some such sub-types can be useful in or with particular methods (generally as described elsewhere herein). Types of riboswitches include, for example, naturally occurring riboswitches, derivatives and modified forms of naturally occurring riboswitches, chimeric riboswitches, and recombinant riboswitches. A naturally occurring riboswitch is a riboswitch having the sequence of a riboswitch as found in nature. Such a naturally occurring riboswitch can be an isolated or recombinant form of the naturally occurring riboswitch as it occurs in nature. That is, the riboswitch has the same primary structure but has been isolated or engineered in a new genetic or nucleic acid context. Chimeric riboswitches can be made up of, for example, part of a riboswitch of any or of a particular class or type of riboswitch and part of a different riboswitch of the same or of any different class or type of riboswitch; part of a riboswitch of any or of a particular class or type of riboswitch and any non-riboswitch sequence or component. Recombinant riboswitches are riboswitches that have been isolated or engineered in a new genetic or nucleic acid context.

[0029] Different classes of riboswitches refer to riboswitches that have the same or similar trigger molecules or riboswitches that have the same or similar overall structure (predicted, determined, or a combination). Riboswitches of the same class generally, but need not, have both the same or similar trigger molecules and the same or similar overall structure.

[0030] Additional advantages of the disclosed method and compositions will be set forth in part in the description which follows, and in part will be understood from the description, or can be learned by practice of the disclosed method and compositions. The advantages of the disclosed method and compositions will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the disclosed method and compositions and together with the description, serve to explain the principles of the disclosed method and compositions.

[0032] FIGS. 1A and 1B show metabolite-dependent conformational changes in the 202-nucleotide leader sequence of the butB mRNA. FIG. 1A shows separation of spontaneous RNA-cleavage products of the butB leader using denaturing 10% polyacrylamide gel electrophoresis (PAGE). 5'-32P-labeled mRNA leader molecules (arrow) were incubated for 41 h at 25°C in 20 mM MgCl₂, 50 mM Tris•HCl (pH 8.3 at 25°C) in the presence (+) or absence (−) of 20 μM of AdoCbl. Lanes containing RNAs that have undergone no reaction, partial digest with alkali, and partial digest with RNase T1 (G-specific cleavage) are identified by NR, “−”OH, and T1, respectively. The location of product bands corresponding to cleavage after selected guanosine residues are identified by filled arrowheads. Arrowheads labeled 1 through 8 identify eight of the nine locations that exhibit effector-induced structure modulation, which experience an increase or decrease in the rate of spontaneous RNA cleavage. The image was generated using a phosphorimager (Molecular Dynamics), and cleavage yields were quantitated by using ImageQuant software. FIG. 1B shows sequence and secondary-structure model for the 202-nucleotide leader sequence of butB mRNA (SEQ ID NO:1) in the presence of AdoCbl. Putative base-paired elements are designated P1 through P9. Complementary nucleotides in the loops of P4 and P9 that have the potential to form a pseudoknot are juxtaposed. Nine specific sites of structure modulation are identified by arrowheads. The asterisks demark the boundaries of the B₃ box (nucleotides 141-162). The coding region and the 38 nucleotides that reside immediately 5' of the start codon (nucleotides 241-243) were not included in the 202-nucleotide fragment. The 315-nucleotide fragment includes the 202-nucleotide fragment, the remaining 38 nucleotides of the leader sequence, and the first 75 nucleotides of the coding region.

[0033] FIGS. 2A and 2B show the butB mRNA leader forms a saturable binding site for AdoCbl. FIG. 2A shows the dependence of spontaneous cleavage of butB3 mRNA leader on the concentration of AdoCbl effector as represented by site 1 (G23) and site 2 (U68). 5'-32P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the in the brief description of FIG. 1, and include identical control and marker lanes as indicated. Incubations contained concentrations of AdoCbl ranging from 10 nM to 100 μM (lanes 1 through 8) or did not include AdoCbl (−). FIG. 2B shows a composite plot of the fraction of RNA cleaved at six locations along the mRNA leader versus the logarithm of the concentration (c) of AdoCbl. Fraction cleaved values were normalized relative to the highest and lowest cleavage values measured for each location, including the values obtained upon incubation in the absence of AdoCbl. The inset defines the symbols used for each of six sites, while the remaining three sites were excluded from the analysis due to weak or obscured cleavage bands. Filled and open symbols represent increasing and decreasing cleavage yields, respectively, upon increasing the concentration of AdoCbl. The dashed line
reflects a K_d of ~300 nM, as predicted by the concentration needed to generate half-maximal structural modulation. Data plotted were derived from a single PAGE analysis, of which two representative sections are depicted in FIG. 1A.

[0034] FIG. 3 shows the 202-nucleotide mRNA leader causes an unequal distribution of AdoCbl in an equilibrium dialysis apparatus. I: Equilibration of tritiated effector was conducted in the absence of RNA. II: (step 1) Equilibration was conducted as in I, but with 200 pmol of mRNA leader added to chamber b; (step 2) 5000 pmol of unlabeled AdoCbl was added to chamber b. III: Equilibrations were conducted as described in II, but wherein 5000 pmol of cyanocobalamin was added to chamber b. IV: (step 1) Equilibration was initiated as described in step 1 of II; (steps 2 and 3) the solution in chamber a was replaced with 25 μL of fresh equilibration buffer; (step 4) 5000 pmol of unlabeled AdoCbl was added to chamber b. The cpm ratio is the ratio of counts detected in chamber b relative to that of a. The dashed line represents a cpm ratio of 1, which is expected if equal distribution of tritium is established.

[0035] FIGS. 4A and 4B show selective molecular recognition of effectors by the butB mRNA leader. FIG. 4A shows a chemical structure of AdoCbl (1) and various effector analogs (2 through 11, ref 30). FIG. 4B shows a determination of analog binding by monitoring modulation of spontaneous cleavage of the 202-nucleotide butB RNA leader. 5'−32P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the legend to FIG. 1G, and include identical control and marker lanes as indicated. The sections of three PAGE analyses encompassing sites 2 (U68) are depicted. Below each image is plotted the amount of RNA cleaved (normalized with reference to the lowest and highest levels of cleavage at U68 in each gel) for each effector as indicated, or for no effector (−). The compound 11 (13-epi-AdoCbl) is the epimer of AdoCbl wherein the configuration at C13 is inverted, so that the e-propionamide side chain is above the plane of the corrin ring; see Brown et al., Conformational studies of 5′-deoxyadenosyl-13-epi-cobalamin, a coenzymatically active structural analog of coenzyme B_{12}. Polyhedron 17, 2213 (1998).

[0036] FIGS. 5A, 5B, SC, SD, SE and SF show mutations in the mRNA leader and their effects on AdoCbl binding and genetic control. FIG. 5A shows sequence of the putative P5 element of the wild-type 202-nucleotide butB leader exhibits AdoCbl-dependent modulation of structure as indicated by the observed increase in spontaneous RNA cleavage at position U68 (10% denaturing PAGE gel). Assays were conducted in the absence (−) or presence (+) of 5 μM of AdoCbl. The remaining lanes are as described in the legend to FIG. 1A. The composite bar graph reflects the ability of the RNA to shift the equilibrium of AdoCbl in an equilibrium dialysis apparatus and the ability of a reporter gene (see Experimental Procedures) to be regulated by AdoCbl addition to a bacterial culture. (Left) Plotted is the cpm ratio derived by equilibrium dialysis, wherein chamber b contains the RNA. Details of the equilibrium dialysis experiments are described in the text of FIG. 3. (Right) Plotted are the expression levels of β-galactosidase as determined from cells grown in the absence (−) or presence (+) of 5 μM AdoCbl. Boxed numbers on the left and right, respectively, reflect the approximate K_d and the fold repression of β-galactosidase activity in the presence of AdoCbl. N.D. designates not determined. FIG. 5B−5F shows sequences and performance characteristics of various mutant leader sequences as indicated. Constructs were created as described in the Experimental Procedures section.

[0037] FIGS. 6A, 6B, 6C and 6D show metabolite binding by mRNAs. FIG. 6A shows TPP-dependent modulation of the spontaneous cleavage of 165 thiM RNA was visualized by polyacrylamide gel electrophoresis (PAGE). 5′−32P-labeled RNAs (arrow, 20 nM) were incubated for approximately 40 hr at 25°C in 20 mM MgCl_2, 50 mM Tris-HCl (pH 8.3) at 25°C in 20 mM MgCl_2, 50 mM Tris-HCl (pH 8.3) at 25°C. The presence (+) or absence (−) of 100 μM TPP, NR, OH and T1 represent RNAs subjected to no reaction, partial digestion with alkali, or partial digestion with RNase T1 (G-specific cleavage), respectively. Product bands representing cleavage after selected G residues are numbered and identified by filled arrowheads. The asterisk identifies modulation of RNA structure involving the Shine-Dalgarno (SD) sequence. Gel separations were analyzed using a phosphorimage system (Molecular Dynamics) and quantitated using ImageQuant software. FIG. 6B shows a secondary-structure model of 165 thiM (SEQ ID NO.2) as predicted by computer modeling (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Biotechnology (eds. Barciszewski J. & Clark, B. F. C.) 11-43 (NATO ASI Series, Kluwer Academic Publishers, 1999); Mathews et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911-940 (1999)) and by the structure probing data depicted in FIG. 6A. Spontaneous cleavage characteristics are as noted in the inset. Unmarked nucleotides exhibit a constant but low level of degradation. The truncated 91 thiM RNA (residues 1−91 of SEQ ID NO.2) is boxed and the box/MB element (Miranda-Rios et al., A conserved RNA structure (boxB) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc. Natl. Acad. Sci. USA 19, 9736-9741 (2001)) is shaded. Nucleotides enclosed in boxes identify an alternative pairing, designated P18'. The RNA carries two mutations (G156A and U157C) relative to wild type that were introduced in a non-essential portion of the construct to form a restriction site for cloning, while all RNAs carry two 5′-terminal G residues to facilitate in vitro transcription. FIG. 6C shows TPP-dependent modulation of the spontaneous cleavage of 240 thiC RNA. Reactions were conducted and analyzed as described in above for FIG. 6A. FIG. 6D shows a secondary-structure model of 240 thiC (SEQ ID NO.3). Base-paired elements that are similar to those of thiM are labeled P1 through P5. The truncated RNA 111 thiC (residues 1-111 of SEQ ID NO.3) is boxed. Nucleotides enclosed in boxes identify an alternative pairing.

[0038] FIGS. 7A, 7B and 7C show the thiM and thiC mRNA leaders serve as high-affinity metabolite receptors. FIG. 7A shows the extent of spontaneous modulation of RNA cleavage at several sites within 165 thiM (left) and 240 thiC (right) plotted for different concentrations (c) of TPP. Arrows reflect the estimated concentration of TPP needed to attain half maximal modulation of RNA (apparent K_d). FIG. 7B shows the logarithm of the apparent K_d values plotted for both RNAs with TPP, TP and thiamine as indicated. The boxed data was generated using TPP with the truncated RNAs 91 thiM and 111 thiC. FIG. 7C shows that patterns of spontaneous cleavage of 165 thiM differ between thiamine and TPP ligands as depicted by PAGE analysis (left) and as reflected by graphs (right) representing the relative phosphoimager counts for the three lanes as indicated. Details for the
RNA probing analysis are similar to those described above in connection with FIG. 6A. The graphs were generated by ImageQuant software.

[0039] FIGS. 8A, 8B, 8C and 8D show high sensitivity and selectivity of mRNA leaders for metabolite binding. FIG. 8A shows chemical structures of several analogues of thiamine. ID is thiamine disulfide and T12 is 4-methyl-5-β-hydroxyethylthiazole. FIG. 8B shows PAGE analysis of 165 thim RNA structure probing using TTP and various chemical analogues (40 μM each) as indicated. Locations of significant structural modulation within the RNA spanning nucleotides -113 to +150 are indicated by open arrowheads. The asterisk identifies the site (C144) used to compare the normalized fraction of RNA that is cleaved (bottom) in the presence of specific compounds. Details for the RNA probing analysis are similar to those described above in connection with FIG. 6A. FIG. 8C shows a summary of the features of TTP that are critical for molecular recognition. FIG. 8D shows equilibrium dialysis using 3H-thiamine as a tracer. Plotted are the ratios for tritium distribution in a two-chamber system (a and b) that were established upon equilibration in the presence of the RNA constructs in chamber b as indicated (see below for a description of the non-TPP-binding mutant M3). 100 μM TPP or oxythiamine were added to chamber a, as denoted, upon the start of equilibration.

[0040] FIGS. 9A, 9B, 9C and 9D show mutational analysis of the structure and function of the thim riboswitch. FIG. 9A shows mutations present in constructs M1 through M8 relative to the 165 thim RNA (SEQ ID NO:4). P8* is a putative base-paired element between portions (encircled) of the P1 and P8 stems. FIGS. 9B and 9C show in vitro ligand-binding and genetic control functions of the wild-type (WT), M1 and M2 RNAs as reflected by PAGE analysis of in-line probing experiments (10 μM TTP) and by β-galactosidase expression assays. Labels on PAGE gels are as described above in connection with FIG. 6A. Bars represent the levels of gene expression in the presence (+) and the absence (−) of TTP in the culture medium. FIG. 9D is a summary of similar analyses of WT through M9 is presented in table form. The SD status “n.d.” (not determined) indicates either that the level of spontaneous cleavage detected in the absence and presence of TTP is near the limit of detection (M6, M7 and M8) or that the region adopts an atypical structure (M9) compared to WT.

[0041] FIG. 10 shows a construct for the selection of SAM-responsive ribozymes (SEQ ID NO:5). The hammerhead self-cleaving ribozyme and the SAM aptamer both require proper formation of the bridge domain to exhibit function. Therefore, the selection is expected to permit ribozyme function only when SAM or another binding-competent analog is present.

[0042] FIGS. 11A (SEQ ID NO:6 and SEQ ID NOs:378-382), 11B (SEQ ID NO:7 and SEQ ID NOs:383-385), 11C (SEQ ID NO:8 and SEQ ID NOs:386-387), 11D (SEQ ID NO:9 and SEQ ID NOs:388-389), 11E (SEQ ID NO:10), 11F (SEQ ID NO:11) and 11G (SEQ ID NO:12 and SEQ ID NOs:390-397) show consensus sequences and putative secondary structures derived by phylogenetic and biochemical analyses as described for each riboswitch (see references). Nucleotides identified by a lower case a, c, t, or g, are conserved in greater than 90% of the representative sequences, open circles identify nucleotide positions of variable sequence, and lines identify elements that are variable in sequence and length. Models are described as follows: 11A) coenzyme B12 aptamer (Example 1); 11B) TTP aptamer (Example 2); 11C) FMN aptamer (Example 3); 11D) SAM aptamer (Example 7); 11E) guanine aptamer (Example 6); 11F) adenine aptamer (Example 8); and 11G) lysine aptamer (Example 5). Letters R and Y represent purine and pyrimidine bases, respectively; K designates G or U; W designates A or U; H designates A, C, or U; D designates G, A, or U; N represents any of the four bases.

[0043] FIGS. 12A (SEQ ID NO:13), 12B and 12C show the regulation of the B. subtilis ribD mRNA by FMN. FIG. 12A shows the results of in-line probing assays. Internucleotide linkages identified with squares exhibit decreased amounts of spontaneous cleavage when ribD is incubated in the presence of FMN (indicating an increase in order for these nucleotides) relative to incubation in the absence of FMN. Circles identify linkages that exhibit consistently high levels of scission, which indicates they are not modulated by presence of FMN. FIG. 12B shows a model for the mechanism of ribD regulation. The ribD mRNA adopts anti-termination conformation in the absence of FMN. Increased levels of FMN stabilize an RFN-FMN complex that permits formation of the terminator structure. FIG. 12C shows the chemical structure and apparent dissociation constants for riboflavin and FMN.

[0044] FIGS. 13A (residues 1-91 of SEQ ID NO:2), 13B and 13C show the regulation of the E. coli thim mRNA by TTP. FIG. 13A shows results of in-line probing assays. Internucleotide linkages identified with squares exhibit decreased amounts of spontaneous cleavage when thim is incubated in the presence of TTP compared to incubation in the absence of ligand. In contrast, linkages identified with hexagons exhibit increased amounts of cleavage when thim is incubated with TTP compared to incubation in the absence of ligand. The boxed nucleotides indicate pyrophosphate-recognition region (as described in text). FIG. 13B shows a model for the mechanism of thim regulation. In the absence of TTP, the anti-SD sequence interacts with part of aptamer domain to form anti-anti-SD. As TTP is increased, aptamer-TTP complexes are formed and the anti-SD favors pairing with the SD. FIG. 13C shows the chemical structure and apparent dissociation constants for thiamine and TTP.

[0045] FIGS. 14A, 14B and 14C show putative eukaryote riboswitches. FIG. 14A shows the consensus TPP binding domain based on 100 bacteria and archaea RNAs (SEQ ID NO:18 and SEQ ID NOs:398-399). Nucleotides shown as lower case letters are most conserved (>90%). Open circles represent nucleotide positions and domains that vary in sequence and length are designated var. The consensus model is similar to that reported recently (Rodionov et al., 2002). FIG. 14B the TPP-binding domain of A. thaliana (SEQ ID NO:14). Variations in O. sativa (nucleotides enclosed in a circle) (SEQ ID NO:15) and P. secunda (nucleotides enclosed in a hexagon) (SEQ ID NO:16) are shown. FIG. 14C shows a putative TPP-binding domain in the intron of N. crassa (SEQ ID NO:17).

[0046] FIG. 15 shows sequence alignments of eukaryotic domains related to bacterial TPP-dependent riboswitches, Eco1, Eco2, Cac1, Ncr1, Aor1, Fox1, Fso1, Ath1, Pse1, Osa1, which are represented by SEQ ID NOs:19-28 respectively. Base paired stems are shaded in black and labeled as defined in Example 2). The P3 sequences, which in eukaryotes are significantly expanded in length and number of base pairs, are represented as a stem-loop structure. The highly conserved nucleotide positions in bacteria that were used to search for eukaryotic domains are enclosed in a box. For each identified (ID) sequence, the position of the conserved CUGAGA
sequence within the given Genbank entry is given along with the accession identification, sequence name, and gene identification. Additional protein annotations based on sequence similarity are shown in brackets. Methods: Riboswitch-like domains were initially identified by sequence similarity to bacterial sequences (Eco2 and CueC) by a blastn search of Genbank using default parameters. These hits were verified and expanded by searching for degenerate matches to the pattern (CTGAGA) [200] AACYGGA [50] C<GT>NNNNNC[>][5] CGNRGGRA (SEQ ID NO: 375). Angle brackets indicate base pairing and bracketed numbers are variable gaps with constrained maximum lengths. All of the eukaryotic sequences have one or zero mismatches to this pattern except for one (Asr) that initially had three mismatches due to a single A insertion in the final search element. This mutation was removed to simplify the alignment. Comparison of mRNA (M33643.1) and genomic (AB033416.1) sequences demonstrated that the F. oxysporum element is in an intron in the 5' UTR of the sti5 gene. Other fungal sequences (Ner, Aor, and Fso) are flanked by consensus splicing sequences.

Fig. 10A and 10B show the structural probing of the putative TPP-riboswitch from Arabidopsis. Fig. 10A shows the fragmentation pattern of the 128-nucleotide RNA (arrow) of A. thaliana (Fig. 10B) which was generated by incubation in the absence (−) or presence (+) of 100 μM TPP, T1, OH and NR identify RNAs that were partially digested with RNase T1 (cleaves 3' to G residues), alkali, or were not reacted, respectively. Reactions were conducted as described in Example 2. Fig. 10B shows the apparent Kp for TPP binding by the A. thaliana RNA. Fraction bound was determined by in-line probing as described in Examples 1-3.

Fig. 11 shows genetic structures thiamine biosynthetic genes and possible mechanisms of riboswitch control. The location and mechanism of the E. coli and B. subtilis riboswitches are detailed in Examples 2 and 6. The putative TPP riboswitch from P. secunda resides immediately upstream from the polyA tail in the cDNA clone of the THIC gene. The putative TPP riboswitch domain in F. oxysporum is located in a 5'-UTR intron of the STI35 gene according to the genomic sequence but is absent in the cDNA clone.

Fig. 12A and 12B show the L box—a highly conserved sequence and structural domain is present in the 5'-UTRs of Gram-positive and Gram-negative bacterial mRNAs that are related to lysine metabolism. Conserved portions of the L box sequence and secondary structure were identified by alignment of 28 representative mRNAs as noted. Base pairing potential representing P1 through P5 are enumerated and set off by boxes. Nucleotides shown as lower case letters are conserved in greater than 80% of the examples. The asterisk identifies the representative (B. subtilis lysC 5'-UTR) that was examined in this study. Gene names are also annotated in GenBank or were derived by protein sequence similarity. Organism abbreviations are as follows: Bacillus anthracis (BA), Bacillus halodurans (BHI), Bacillus subtilis (BS), Clostridium acetobutylicum (CA), Clostridium perfringens (CP), Escherichia coli (EC), Haemophilus influenzae (HI), Oceanobacillus iheyensis (OI), Pasteurella multocida (PM), Staphylococcus aureus (SA), Staphylococcus epidermidis (SE), Shigella flexneri (SF), Shewanella oneidensis (SO), Thermatoga maritima (TM), Thermoaerobacter tengcongensis (TT), Vibrio cholerae (VC), Vibrio vulnificus (VV), Thermoaerobacter tengcongensis (TT).

Fig. 11A (SEQ ID NO:60 and SEQ ID NO:61) show the consensus L box motif from the lysC 5'-UTR of B. subtilis undergoes allosteric rearrangement in the presence of L-lysine. (A) Consensus sequence and structure of the L box domain as derived using a phylogeny of 31 representative sequences from prokaryotic and archael organisms (Fig. 10B) BA 0845, BA lysA, BA lysP, BH dapa, BH lysC, BH nhaC, BS lysC, BX lysC, CA lysA, CP lysA, CP lysP, EC lysC, HI nhaC, OI dapa, OI nhaC, PM nhaC, SA lysC, SA lysP, SE lysC, SE lysP, SF lysC, SO lysC, SO nhaC, TM asd, TT lysA, TT pspF, VC lysC, VC nhaC, VC nhaC, VV lysC, VV nhaC, which are represented by SEQ ID NOs:29-59, respectively. Nucleotides depicted a lower case a, c, t, or g, are present in at least 80% of the representatives, open circles identify nucleotide positions of variable identity, and dashed lines denote variable nucleotide identity and chain length. Fig. 11B shows sequence, secondary structure model, and lysine-induced structural modulation of the lysC 5'-UTR of B. subtilis. An additional 94 nucleotides (not depicted) reside between nucleotide 237 and the AUG start codon. Structural modulation sites (nucleotides enclosed in squares) were established using 237 lysC RNA by monitoring spontaneous RNA cleavage as depicted in C. Fig. 11C shows in-line probing of the 237 lysC RNA reveals lysine-induced modulation of RNA structure. Patterns of spontaneous cleavage, revealed by product separation using denaturing 10% polyacrylamide gel electrophoresis (PAGE), are altered at four major sites (denoted 1 through 4) in the presence (+) of 10 μM L-lysine (L) relative to that observed in the absence (−) of lysine. T1, OH and NR represent partial digest with RNase T1, partial digest with alkali, and no reaction, respectively. Selected bands in the T1 lane (G-specific cleavage) are identified by nucleotide position. See Methods for experimental details.

Fig. 12A, 20B, 20C, 20D and 20E show the molecular recognition characteristics of the lysine aptamer and the use of caged lysine. Fig. 20A shows the chemical structures of L-lysine, D-lysine and nine closely-related analogs. Small circles represent chiral carbon centers wherein the enantiomeric configuration is defined for each compound. Encircled atoms identify chemical differences between L-lysine and the analog depicted. Fig. 20B shows in-line probing analysis of the 179 lysC RNA in the absence (−) of ligand, or in the presence of 10 μM L-lysine or 100 μM of various analogs as indicated for each lane. For each lane, the relative extent of spontaneous cleavage at site 3 is compared to that of the zone of constant cleavage immediately below this site, where a cleavage ratio significantly below ~1.5 reflects modulation. Fig. 20C shows a schematic representation of dipeptide digestion by hydrochloric acid. All dipeptide forms are expected to be incapable of binding the lysine aptamer (inactive), while lysine-containing dipeptides should induce conformational changes in the aptamer (active) upon acid digestion. Fig. 20D shows in-line probing analysis of the 179 lysC RNA in the absence of lysine (−) or in the presence of various amino acids and dipeptides. Underlined lines carry dipeptide preparations that were pretreated with HCl as depicted in a. Fig. 20E shows the fraction of spontaneous cleavage at site 3 in d is plotted after normalization to the extent of processing in the absence of added ligand.

Fig. 13A, 21B, 21C and 21D show determination of the dissociation constant and stoichiometry for L-lysine binding to the 179 lysC RNA. Fig. 21A shows in-line probing with increasing concentrations of L-lysine ranging from 3 nM.
to 3 mM. Details are as defined for FIG. 19C. FIG. 20B shows a plot depicting the normalized fraction of RNA undergoing spontaneous cleavage versus the concentration of amino acid for sites 1 through 3. The dashed line identifies the concentration of L-lysine required to bring about half-maximal structural modification, which indicates the apparent $K_d$ for ligand binding. FIG. 20C shows the 179 lysC RNA (10 mM) shifts the equilibrium of tritiated L-lysine (50 mM) in an equilibrium dialysis chamber. To investigate competitive binding, unlabeled L-(L) and D-lysine (D), or L-citrulline (S) were added to a final concentration of 50 mM each to one chamber of a pre-equilibrated assay as indicated. FIG. 21D shows a scatchard analysis of L-lysine binding by the 179 lysC RNA. The variable $r$ represents the ratio of bound ligand concentration versus the total RNA concentration and the variable $[L]_0$ represents the concentration of free ligand.

[0053] FIGS. 22A, 22B and 22C show the B. subtilis lysC riboswitch and its mechanism for metabolite-induced transcription termination. FIG. 22A shows a sequence and repressed-state model for the lysC riboswitch secondary structure (SEQ ID NO:62). The encircled nucleotide indicates the putative anti-terminator interaction that could form in the absence of L-lysine. Boxed nucleotides identify sites of disruption (M1) and compensatory mutations for the terminator C (M2) and for the terminator and anti-terminator stems (M3). Nucleotides enclosed in squares identify some of the positions where mutations exhibit lysC derepression that were reported previously (Vold et al. 1975; Lu et al. 1992). FIG. 22B shows in vitro transcription assays conducted in the absence (-) or presence (+) of 10 mM L-lysine or other analogs as indicated. FL and F identify the full-length and terminated transcripts, respectively. The percent of the terminated RNAs relative to the total terminated and full-length transcripts are provided for each lane (% term.). FIG. 22C shows in vivo expression of a β-galactosidase reporter gene fused to wild-type (WT), G39A and G40A mutant lysC 5’S-UTR fragments. Media conditions are as follows: I, normal medium (0.27 mM lysine); II, minimal medium (0.012 mM); III, lysine-supplemented minimal medium (1 mM); IV, lysine hydroxamate-supplemented (medium II plus 1 mM lysine hydroxamate) minimal media; VI, thiosulfate-supplemented (medium II plus 1 mM thiosulfate) minimal medium.

[0054] FIG. 23 depicts a highly conserved domain in the 5’S-UTR of certain gram-positive and gram-negative bacterial mRNAs. Depicted is an alignment of 32 representative mRNA domains from bacteria that conform to the G box consensus sequence (BH1-guaA, BH2-[phuG], BH3-[padR], BH4-snaA, BH5-xpt), BS1-[phuG], BS2-[padR], BS3-xpt, BS4-xpt, BS5-xpt, BS6-xpt, CA1-xpt, CA2-[phuG], CA3-[guaB], CP1-xpt, CP2-aptC, CP3-[guaB], CP4-aptG, FN1-aptQ, LL1-xpt, LM1-[phuG], LM2-[xpt], O11-guaA, O12-[phuG], O13-aptC, O14-aptC, SA1-xpt, TSE1-[xpt], STA1-xpt, STPY1-xpt, STPN-xpt, TE1-[phuG], VV1-add, which are represented by SEQ ID NO:63-94 respectively. Enclosed and enumerated regions identify base-pairing potential of stems P1, P2, and P3, respectively. Nucleotides shown as lower case letters are conserved in greater than 90% of the examples. The asterisk identifies the representative (xpt-phuX 5’S-UTR) that was examined in this study. It is important to note that three representives (BS5, CP4 and VV1) that carry a C to U mutation in the conserved core (in the P3-P1 junction) appear to be adenine-specific riboswitches (unpublished observations). Gene names are as annotated in GenBank, the SubtiList database, or based on protein similarity searches (brackets). Organisms abbreviations are as follows: Bacillus halodurans (BH), Bacillus subtilis (BS), Clostridium acetobutylicum (CA), Clostridium perfringens (CP), Fusobacterium nucleatum (FN), Lactococcus lactis (LL), Listeria monocytogenes (LM), Oceanobacillus iheyensis (OI), Staphylococcus aureus (SA), Staphylococcus epidermidis (SE), Streptococcus agalactiae (STA), Streptococcus pyogenes (STP), Streptococcus pneumoniae (STP), Thermus thermophilus (T), and Fibro-Vulviflu-cus (VV).

[0055] FIGS. 24A, 24B and 24C show the G box RNA of the xpt-phuX mRNA in B. subtilis responds allosterically to guanine. FIG. 24A shows the consensus sequence and secondary model for the G box RNA domain that resides in the 5’S-UTR of genes that are largely involved in purine metabolism (SEQ ID NO:95). Phylogenetic analysis is consistent with the formation of a three-stem (P1 through P3) junction. Nucleotides depicted as shown as lower case letters and capitals are present in greater than 90% and 80% of the representatives examined, respectively (FIG. 23). Encircled nucleotides exhibit base complementation, which might indicate the formation of a pseudoknot. FIG. 25B shows sequence and ligand-induced structural alterations of the 5’S-UTR of the xpt-phuX transcriptional unit (SEQ ID NO:96). The putative anti-terminator interaction is represented by the boxes. Nucleotides that undergo structural alteration as determined by in-line probing (from C) are identified with squares. The 93 xpt fragment (boxed) of the 201 xpt RNA retains guanine-binding function. Asterisks denote alterations to the RNA sequence that facilitate in vitro transcription (5’S terminus) or that generate a restriction site (3’S terminus). Nucleotide numbers begin at the first nucleotide of the natural transcription start site. The translation start codon begins at position 186. FIG. 24C shows guanine and related purines selectively induce structural modification of the 93 xpt mRNA fragment. Precursor RNAs (Pre; [5-32P]-labeled) were subjected to in-line probing by incubation for 40 hr in the absence (-) or presence of guanine, hypoxanthine, xanthine and adenine as indicated by G, H, X and A, respectively. Lanes designated NR, T1 and 5’S-UTR contain RNA that was not reacted, subjected to partial digestion with RNase T1 (G-specific cleavage), or subjected to partial alkaline digestion, respectively. Selected bands corresponding to G-specific cleavage are identified. Regions 1 through 4 identify major sites of ligand-induced modulation of spontaneous RNA cleavage. FIGS. 25A and 25B show the 201 xpt mRNA Leader Binds Guanine with High Affinity. FIG. 25A shows in-line probing reveals that spontaneous RNA cleavage of the 201 xpt RNA at four regions decreases with increasing guanine concentrations. Only those locations of the PAGE image corresponding to the four regions of modulation as indicated in FIG. 25C are depicted. Other details and notations are as described in the legend to FIG. 25C. FIG. 25B shows a plot depicting the normalized fraction of RNA that experienced spontaneous cleavage versus the concentration of guanine for modulated regions 1 through 4 in FIG. 25A. Fraction cleaved values were normalized to the maximum cleavage measured in the absence of guanine and to the minimum cleavage measured in the presence of 10 μM guanine. The apparent $K_d$ value (less than or equal to 5 nM) reflects the limits of detection for these assay conditions.

[0057] FIGS. 26A, 26B and 26C show a molecular discrimination by the guanine-binding aptamer of the xpt-phuX mRNA. FIG. 26A shows the chemical structures and apparent
K_p values for guanine, hypoxanthine and xanthine (active natural regulators of xpt-phbX genetic expression in B. subtilis) versus that of adenine (inactive). Differences in chemical structure relative to guanine are encircled. K_p values were established as shown in FIG. 26 with the 201 xpt RNA. Numbers on guanine represent the positions of the ring nitrogen atoms. FIG. 26B shows chemical structures and K_p values for various analogs of guanine reveal that all alterations of this purine cause a loss of bonding affinity. Open circles identify K_p values that most likely are significantly higher than indicated, as concentrations of analog above 500 μM were not examined in this analysis. The apparent K_p values of G, H, X and A as indicated are plotted as triangles for comparison. FIG. 26C shows a schematic representation of the molecular recognition features of the guanine aptamer in 201 xpt. Hydrogen bond formation at position 9 of guanine is expected because guanosine (K_p >100 μM) and inosine (K_p >100 μM), which are N-ribosyl derivatives of guanine and hypoxanthine, respectively, do not exhibit measurable binding (see FIG. 27).

[0058] FIGS. 27A and 27B show confirmation of guanine binding specificity by equilibrium dialysis. FIG. 27A shows an equilibrium dialysis strategy was used to confirm that in vitro-transcribed 93 xpt RNAs bind to guanine and can discriminate against various analogs. Each data point was generated by adding 3H-guanine to chamber a, which is separated from RNA and other analogs by a dialysis membrane with a molecular weight cut-off of 5,000 daltons. Left: If no guanine binding sites are present in chamber b, or if an excess of unlabeled competitor is present, then no shift in the distribution of tritium is expected. Right: If an excess of guanine-binding RNAs are present in chamber b, and if no competitor is present, then a substantial shift in the distribution of tritium towards chamber b is expected. FIG. 27B shows the 93 xpt RNA can shift the distribution of 3H-guanine in an equilibrium dialysis apparatus, while analogs of guanine are poor competitors. The plot depicts the fraction of counts per minute (cpm) of tritium in chamber b relative to the total amount of cpm counted from both chambers. A value of ~0.5 is expected if no shift occurs, as is the case when RNA is absent (none), or in the presence of excess unlabeled competitor (G). A value approaching 1 is expected if the majority of 3H-guanine is bound by the RNA in chamber b in the absence of unlabeled analog, or in the presence of unlabeled analogs that do not serve as effective competitors under the assay conditions (100 nM 3H-guanine, 300 nM RNA, 500 nM analog). Ino and Gua represents inosine and guanosine, respectively.

[0059] FIGS. 28A, 28B, 28C and 28D show the binding and genetic control functions of variant guanine riboswitches. FIG. 28A shows mutations used to examine the importance of various structural features of the guanine aptamer domain (SEQ ID NO:97). FIG. 28B shows examination of the binding function of aptamer variants by equilibrium dialysis. WT designates the wild-type 93 xpt construct. Details are as described for FIG. 27. FIG. 28C shows genetic modulation of a β-galactosidase reporter gene upon the introduction of various purines as indicated. FIG. 28D shows regulation of β-galactosidase reporter gene expression by WT and mutants M1 through M7. Open and filled bars represent enzyme activity generated when growing cells in the absence and presence of guanine, respectively.

[0060] FIGS. 29A, 29B and 29C show that riboswitches participate in fundamental genetic control. FIGS. 29A and 29B are schematic representations of the seven known riboswitches and the metabolites they sense. The secondary structure models were obtained as follows: coenzyme B_12 (see Example 1); TPP (see Example 2); FMN (see Example 3); SAM (see Example 7); guanine (see Example 6); lysine (see Example 5); adenine (see Example 8). Coenzyme B_12 is depicted in exploded form wherein a, b and c designate covalent attachment sites between fragments. FIG. 29C shows a genetic map of B. subtilis riboswitch regulations and their positions on the bacterial chromosome. Genes are controlled by riboswitches as identified by matching numbers. All nomenclature is derived from the SubtiList database release R16.1 (Moszer, I., et al., 1995, Microbiol. 141, 261-268) except for metII and metC, which are recent designations (Auger, S., et al., 2002, Microbiol. 148, 507-518).

[0061] FIGS. 30A, 30B and 30C show the S Box is a structured RNA domain that binds SAM. (A) Consensus sequence and secondary-structure model of the S box domain derived from 107 bacterial representatives (SEQ ID NO:98 and SEQ ID NO:99). Lower case letter and capital letter positions identify nucleotides whose identity as depicted is conserved in greater than 90% or 80% of the representative S box RNAs, respectively. R, Y, and N represent purine, pyrimidine, and any nucleotide, respectively. P1 through P4 identify conserved base pairing. Enclosed nucleotides identify a putative pseudoknot interaction. FIG. 30B shows a sequence and secondary structure model for the 251 yitJ mRNA fragment (SEQ ID NO:99). Sites of structural modulation upon introduction of SAM are depicted as described. Nucleotide 1 corresponds to the putative transcriptional start site. Asterisks identify nucleotides that were added to the construct to permit efficient transcription in vitro. The first nucleotide of the AUG start codon is 121 (not shown). Other notations are as described in a. FIG. 30C shows the spontaneous cleavage patterns of 251 yitJ (~10 nM 5'-32P-labeled) RNA incubated for ~40 hr at 25°C in 50 nM Tris-HCl (pH 8.3) at 25°C, 20 mM MgCl_2, 100 mM KCl, and without (+) or with methionine or SAM as indicated for each lane. Nt, T1 and T2 OH represent no reaction, partial digest with RNase T1, and partial digest with alkali, respectively. Certain fragment bands corresponding to T1 digestion (cleaves after G residues) are depicted. Arrowheads identify positions of significant modulation of spontaneous cleavage, and the numbered sites were used for quantitation (see FIG. 31B). Experimental procedures are similar to those described in Examples 1-3.

[0062] FIGS. 31A, 31B and 31C show the binding affinity and molecular discrimination by a SAM-binding RNA. FIG. 31A shows the chemical structures of various compounds used to probe the binding characteristics of the SAM yitJ riboswitch. Other than methionine, each compound as depicted is coupled to an adenosyl moiety ([A]; inset) coupled via the 5'-carbon (as signified by R). FIG. 31B: Left: The K_p of 251 yitJ for SAM was determined by plotting the normalized fraction of RNA cleaved at regions 1 through 6 (see FIG. 30C) versus the logarithm of the concentration of SAM in molar units. The dashed line indicates the concentration needed to induce half maximal modulation of cleavage activity. Right: K_p values for SAM and various analogs as determined by this method. FIG. 31C shows molecular discrimination determined by equilibrium dialysis. Assays employed 100 nM of S-adenosyl-L-methionine-methyl-3H (3H-SAM; 14.5 μCi mmol⁻¹; ~7,000 cpm) added to side A of an equilibrium dialysis chamber (1, 2), and were conducted in the absence (none) or the presence of 3 μM RNA on the B side of the chamber as indicated. Equilibrations were carried out
for ~10 hr in the absence (−) of unlabelled analogs, and then were subsequently incubated in the presence of 25 μM unla-
beled compounds (added to side B) as indicated. M1 is a variant of 124 yfβ that carries disruptive mutations in the
junction between stems P1 and P2 (FIG. 32a). Line at a cpm
ratio of 1 identifies the bar height expected if a shift in
3H-SAM has not occurred. Additional experimental details
are similar to those described in Examples 1 and 2.

[0063] FIGS. 32A, 32B and 32C show the effects of RNA
mutations on SAM binding and genetic control. FIG. 32A
shows the sequence and secondary structure model for the
124 yfβ RNA (SEQ ID NO:100). Mutations M1 through M9
were generated in plasmids containing fusions of the yfβ1
5’UTR upstream of a lacZ reporter gene. Templates for
preparation of mutant RNAs for in vitro studies were then
created by PCR, and the mutant DNA constructs were
integrated into the chromosome for in vivo studies. See Methods
for experimental details. FIG. 32B shows the analysis of
SAM-binding function by equilibrium dialysis in the pres-
ence of wild-type (WT) and mutant RNAs as denoted. Details
are described in the legend to FIG. 31C, except that 300 nM
RNA was used and all assays were conducted without the
addition of unlabelled analogs. FIG. 32C shows In vivo control
of β-galactosidase expression in B. subtilis cells trans-
formed with various riboswitch constructs as indicated. β-ga-
lactosidase activities were measured as described in Ex-
ample 2. Cells were grown in glucose minimal media in 0.75 μg
mL−1 1-methionine (+)/50 μg mL−1 1-methionine (+). M6 through
M9 were not examined in vivo.

[0064] FIGS. 33A, 33B, 33C and 33D show metabolite-
induced transcription termination of several mRNAs that
carry a SAM riboswitch. FIG. 33A shows In vitro transcrip-
tion using T7 RNA polymerase results in increased termina-
tion of four mRNA leader sequences. Reactions were
conducted in the absence (−) or presence (+) of 50 μM of the
effector as indicated for each lane. For example, the metT
template includes the 5’ UTR and coding sequences through
mRNA position 242, while the termination site is expected to
occur at position 189. Below each gel is indicated the per-
centage of transcription termination (T) at the expected loca-
tion relative to the total amount of expected termination plus
full length RNA (FL). FIGS. 33B-33D show sequence and struc-
tural model for the metT riboswitch in two structural states
(SEQ ID NO:101). Residues shown in hexagons and squares
correspond to the P1 (anti-anti-terminator) and the terminator
stems, respectively. The encircled residues correspond to the
anti-terminator stem. Sequences boxed in black define the
location and identity of mutations used to examine the pro-
posed mechanism of genetic control. Gel: Analysis of mutant
metT riboswitches wherein disruptive (Ma, Mab and Mc) or
the corresponding compensatory mutations (Mabc) have
been inserted. The metT mutant templates and wild-type con-
trol template (WT) are identical to the templates used in A,
except that the FL product is 220 nucleotides. Other notations
are as described in A.

[0065] FIGS. 34A and 34B show Bacilli species subtilis
and anthrasis bind SAM with different affinities. FIG. 34A
shows structural modulation of the B. subtilis css1 aptamer
as determined by in-line probing (SEQ ID NO:102). Inset:
Apparent K D values determined by monitoring structural
modulation over a range of SAM or SAM analog concen-
trations. Two G residues (asterisks) were included at the 5’
terminus of the RNA construct to facilitate in vitro transcrip-
tion. Nucleotide numbers are given relative to the putative
transcription start site. In-line probing was conducted with an
RNA extending to nucleotide 117, while the remainder of the
RNA is shown to depict the putative transcription terminator
stem. Experiments were similar to those described in FIG.
30B and FIG. 31B. See the legend for FIG. 30B for details.
FIG. 34B shows structural modulation of the B. subtilis css1
aptamer as determined by in-line probing (SEQ ID NO:103).
The transcript start point of the B. anthracis css1 mRNA
has not been determined, and so numbering of nucleotides
begins immediately after the two inserted G residues (aster-
isks). In-line probing was conducted with an RNA extending
to nucleotide 112. See A for additional details.

[0066] FIGS. 35A, 35B and 35C show guanine- and
adenine-specific riboswitches. FIG. 35A shows sequence
and structural features of the two guanine-specific (purL) and
the three adenine-specific aptamer domains that are exam-
ined in this study BS2-purF, BS3-xpt, BS5-ylhL, CP4-add,
VV1-add, which are represented by SEQ ID NOs:104-108,
respectively. P1 through P3 identify the three base-paired
stems comprising the secondary structure of the aptamer
domain. Lowercase nucleotides identify positions whose
base identity is conserved in greater than 90% of representa-
ts in the phylogeny 1. The arrow identifies a nucleotide
within the conserved core of the aptamer that is a determinant
of ligand specificity. BS, CP and VV designate B. subtilis,
Clostridium perfringens and Vibrio vulnificus, respectively.
FIG. 35B shows sequence and secondary structure of the xpt
and ylhL aptamers (SEQ ID NO:109). Encircled nucleotides
identify positions within the ylhL aptamer that differ from
those in the xpt aptamer. The sequence disclosed in FIG. 35C
is SEQ ID NO:110. Nucleotides in xpt are numbered as
described in Example 6. Other notations are as described in A.

[0067] FIGS. 36A, 36B, 36C, 36D and 36E show the ligand
specificity of five G box RNAs. (A through E) In-line probing
assays for the conserved aptamer domains as labeled. NR, T1,
and “OH identify marker lanes wherein precursor RNAs (Pre)
were not incubated, or were partially digested with RNase T1
or alkali, respectively. Selected bands corresponding to
RNase T1 digestion (cleavage 3’ relative to guanylimid residues)
are labeled for each RNA. RNAs were incubated for 40 hr in
the absence of ligand (−), or in the presence of 1 μM guanine
(G) or adenine (A). Large arrowheads identify sites of sub-
stantial change in cleavage pattern that is due to the addition
of a particular ligand. See Methods for additional details.

[0068] FIGS. 37A and 37B show the binding affinity of the
ylhL aptamer for adenine. FIG. 37A shows the in-line probing
assay for the 80 ylhL RNA at various concentrations of
adenine. For each lane, sites 1 through 4 were quantitated
and the fraction of RNA cleaved was used to determine the
apparent K D. FIG. 37B shows a plot of the normalized fraction
of RNA that has undergone spontaneous cleavage at sites 1
through 4 versus the concentration of adenine. See Example
8 for additional details.

[0069] FIGS. 38A and 38B show the specificity of molecu-
lar recognition by the adenine aptamer from ylhL. FIG. 38A
Top: Chemical structures of adenine, guanine and other
purine analogs that exhibit measurable binding to the 80 ylhL
RNA. Chemical changes relative to 2,6-DAP, which is the
tightest-binding compound, are encircled. Bottom left: Plot of
the apparent K D values for various purines. Bottom right:
Model for the chemical features on adenine that serve as
molecular recognition contacts for ylhL. Note that the impor-
ance of N7 and N9 has not been determined. Encircled arrow
indicated that a contact could exist if a hydrogen bond donor

As disclosed herein, distinct classes of riboswitches have been identified and are shown to selectively recognize activating compounds (referred to herein as trigger molecules). For example, coenzyme B$_{12}$, thiamine pyrophosphate (TPP), and flavin mononucleotide (FMN) activate riboswitches present in genes encoding key enzymes in metabolic or transport pathways of these compounds. The aptamer domain of each riboswitch class forms to a highly conserved consensus sequence and structure. Thus, sequence homology searches can be used to identify related riboswitch domains. Riboswitch domains have been discovered in various organisms from bacteria, archaea, and eukarya.

One class of riboswitches that recognizes guanine and discriminates against most other purine analogs has been discovered. Representative RNAs that carry the consensus sequence and structural features of guanine riboswitches are located in the 5'-untranslated region (UTR) of numerous genes of prokaryotes, where they control expression of proteins involved in purine salvage and biosynthesis. Three representatives of this phylegenetic collection bind adenine with values for apparent dissociation constant (apparent K_d) that are several orders of magnitude better than for guanine. The preference for adenine is due to a single nucleotide substitution in the core of the riboswitch, wherein each representative most likely recognizes its corresponding ligand by forming a Watson/Crick base pair. In addition, the adenine-specific riboswitch associated with the ydhL gene of Bacillus subtilis functions as a genetic 'ON' switch, wherein adenine binding causes a structural rearrangement that precludes formation of an intrinsic transcription terminator stem. Guanine-sensing riboswitches are a class of RNA genetic control elements that modulate gene expression in response to changing concentrations of this compound.

It was discovered that the 5'-untranslated sequence of the Escherichia coli thiM mRNA assumes a more protractive role in metabolic monitoring and genetic control. The mRNA serves as a metabolite-sensing genetic switch by selectively binding coenzyme B$_{12}$ without the need for proteins. This binding event establishes a distinct RNA structure that is likely to be responsible for inhibition of ribosome binding and consequent reduction in synthesis of the cobalamin transport protein BtuB. This discovery, along with related observations described herein, supports the hypothesis that metabolic monitoring through RNA-metabolite interactions is a widespread mechanism of genetic control.

RNA structure probing data indicate that the thiamine pyrophosphate (TPP) riboswitch operates as an allosteric sensor of its target compound, wherein binding of TPP by the aptamer domain stabilizes a conformational state within the aptamer and within the neighboring expression platform that precludes translation. The diversity of expression platforms appears to be expansive. The thiM RNA uses a Shine-Dalgarino (SD)-blocking mechanism to control translation. In contrast, the thC RNA controls gene expression both at transcription and translation, and therefore might make use of a somewhat more complex expression platform that converts the TPP binding event into a transcription termination event and into inhibition of translation of completed mRNAs.

A. General Organization of Riboswitch RNAs

Bacterial riboswitch RNAs are genetic control elements that are located primarily within the 5'-untranslated region (5'-UTR) of the main coding region of a particular mRNA. Structural probing studies (discussed further below) reveal that riboswitch elements are generally composed of two domains: a natural aptamer (T. Hermann, D. J. Patel, Science 2000, 287, 820; L. Gold, et al., Annual Review of Biochemistry 1995, 64, 763) that serves as the ligand-binding domain, and an 'expression platform' that interfaces with RNA elements that are involved in gene expression (e.g. Shine-Dalgarino (SD) elements; transcription terminator stems). These conclusions are drawn from the observation that aptamer domains synthesized in vitro bind the appropriate ligand in the absence of the expression platform (see Examples 2, 3 and 6). Moreover, structural probing investigations suggest that the aptamer domain of most riboswitches adopts a particular secondary- and tertiary-structure fold when examined independently, that is essentially identical to the aptamer structure when examined in the context of the entire 5' leader RNA. This implies that, in many cases, the aptamer domain is a modular unit that folds independently of the expression platform (see Examples 2, 3 and 6).

Ultimately, the ligand-bound or unbound status of the aptamer domain is interpreted through the expression platform, which is responsible for exerting an influence upon gene expression. The view of a riboswitch as a modular element is further supported by the fact that aptamer domains are highly conserved amongst various organisms (and even between kingdoms as is observed for the TPP riboswitch), (N. Sudarsan, et al., RNA 2003, 9, 644) whereas the expression platform varies in sequence, structure, and in the mechanism by which expression of the appended open reading frame is controlled. For example, ligand binding to the TPP riboswitch of the thiA mRNA of B. subtilis causes transcription termination (A. S. Mironov, et al., Cell 2002, 111, 747). This expression platform is distinct in sequence and structure compared to the expression platform of the TPP riboswitch in the thiM mRNA from E. coli, wherein TPP binding causes inhibition of translation by a SD blocking mechanism (see Example 2). The TPP aptamer domain is easily recognizable and of near identical functional character between these two transcriptional units, but the genetic control mechanisms and the expression platforms that carry them out are very different.

Aptamer domains for riboswitch RNAs typically range from ~70 to 170 nt in length (FIG. 11). This observation was somewhat unexpected given that in vitro evolution experiments identified a wide variety of small molecule-binding aptamers, which are considerably shorter in length and structural intricacy (T. Hermann, D. J. Patel, Science 2000, 287, 820; L. Gold, et al., Annual Review of Biochemistry 1995, 64, 763; M. Famulok, Current Opinion in Structural Biology 1999, 9, 324). Although the reasons for the substantial increase in complexity and information content of the natural aptamer sequences relative to artificial aptamers remains to be proven, this complexity is most likely required to form RNA receptors that function with high affinity and selectivity. Apparent K_d values for the ligand-riboswitch complexes range from low nanomolar to low micromolar. It is also worth noting that some aptamer domains, when isolated from the appended expression platform, exhibit improved affinity for the target ligand over that of the intact riboswitch. (~10 to 100-fold) (see Example 2). Presumably, there is an energetic cost in sampling the multiple distinct RNA conformations required by a fully intact riboswitch RNA, which is reflected by a loss in ligand affinity. Since the aptamer domain must serve as a molecular switch, this might also add to the
functional demands on natural aptamers that might help rationalize their more sophisticated structures.

B. Riboswitch Regulation of Transcription Termination in Bacteria

[0083] Bacteria primarily make use of two methods for termination of transcription. Certain genes incorporate a termination signal that is dependent upon the Rho protein. (J. P. Richardson, Biochimica et Biophysica Acta 2002, 1577, 251). while others make use of Rho-independent terminators (intrinsic terminators) to destabilize the transcription elongation complex (I. Gusarov, E. Nucler, Molecular Cell 1999, 3, 495; E. Nucler, M. E. Gottesman, Genes to Cells 2002, 7, 755). The latter RNA elements are composed of a GC-rich stem-loop followed by a stretch of 6-9 uridy1 residues. Intrinsic terminators are widespread throughout bacterial genomes (E. Lillo, et al., 2002, 18, 971), and are typically located at the 3’-termini of genes or operons. Interestingly, an increasing number of examples are being observed for intrinsic terminators located within 5’-UTRs.

[0084] Amongst the wide variety of genetic regulatory strategies employed by bacteria there is a growing class of examples wherein RNA polymerase responds to a termination signal within the 5’-UTR in a regulated fashion (T. M. Henkin, Current Opinion in Microbiology 2000, 3, 149). During certain conditions the RNA polymerase complex is directed by external signals either to perceive or to ignore the termination signal. Although transcription initiation might occur without regulation, control over mRNA synthesis (and of gene expression) is ultimately dictated by regulation of the intrinsic terminator. Presumably, one of at least two mutually exclusive mRNA conformations results in the formation or disruption of the RNA structure that signals transcription termination. A trans-acting factor, which in some instances is a RNA (F. J. Grundy, et al., Proceedings of the National Academy of Sciences of the United States of America 2002, 99, 11121; T. M. Henkin, C. Yanofsky, Bioessays 2002, 24, 700) and in others is a protein (J. Stulke, Archives of Microbiology 2002, 177, 433), is generally required for receiving a particular intracellular signal and subsequently stabilizing one of the RNA conformations. Riboswitches offer a direct link between RNA structure modulation and the metabolite signals that are interpreted by the genetic control machinery. A brief overview of the FMN riboswitch from a B. subtilis mRNA is provided below to illustrate this mechanism.

[0085] It was discovered that certain mRNAs involved in thiamine biosynthesis bind to thiamine (vitamin B1) or its bioactive pyrophosphate derivative (TPP) without participation of protein factors. The mRNA-effector complex adopts a distinct structure that sequesters the ribosome-binding site and leads to a reduction in gene expression. This metabolite-sensing mRNA system provides an example of a genetic “riboswitch” (referred to herein as a riboswitch) whose origin might predate the evolutionary emergence of proteins. It has been discovered that the mRNA leader sequence of the btuB gene of Escherichia coli can bind coenzyme B12 selectively, and that this binding event brings about a structural change in the RNA that is important for genetic control (see Example 1). It was also discovered that mRNAs that encode thiamine biosynthetic proteins also employ a riboswitch mechanism (see Example 2).

[0086] It was also discovered that the 5’-UTR of the lysC gene of Bacillus subtilis carries a conserved RNA element that serves as a lysine-responsive riboswitch. The ligand-binding domain of the riboswitch binds to L-lysine with an apparent dissociation constant (Kd) of approximately 1 μM, and exhibits a high level of molecular discrimination against closely related analogs including D-lysine and ornithine. This widespread class of riboswitches serves as a target for the antimicrobial agent thiosine.

[0087] It was also discovered that the xpt-phxX operon (Christiansen, L. C., et al., 1997, J. Bacteriol. 179, 2540-2550) is controlled by a riboswitch that exhibits high affinity and high selectivity for guanine. This class of riboswitches is present in the 5’-untranslated region (5’-UTR) of five transcriptional units in B. subtilis, including that of the 12-gene pur operon. Direct binding of guanine by mRNAs serves as a critical determinant of metabolic homeostasis for purine metabolism in certain bacteria. Furthermore, the discovered classes of riboswitches, which respond to seven distinct target molecules, control at least 68 genes in Bacillus subtilis that are of fundamental importance to central metabolic pathways.

[0088] It was discovered that a highly conserved RNA domain termed the S box serves as a selective aptamer for SAM. Allosteric modulation of secondary and tertiary structures are induced upon SAM binding to the aptamer domain, and these structural changes are responsible for inducing termination of mRNA transcription.

[0089] A variant class of riboswitches that responds to adenosine is also disclosed. These riboswitches carry an aptamer domain that corresponds closely in sequence and secondary structure to the guanine aptamer. However, each representative of the adenosine sub-class of riboswitches carries a C to U mutation in the conserved core of the aptamer, indicating that this residue is involved in metabolic recognition. The identity of this single nucleotide determines the binding specificity between guanine and adenosine, which provides an example of how complex riboswitch structures can be mutated to recognize new metabolite targets.

[0090] Although the specific natural riboswitches disclosed herein are the first examples of mRNA elements that control genetic expression by metabolite binding, it is expected that this genetic control strategy is widespread in biology. It has been suggested (White III, Coenzymes as fossils of an earlier metabolic state. J. Mol. Evol. 7, 101-104 (1976) White III, In: The Pyridine Nucleotide Coenzymes. Acad. Press, NY pp. 1-17 (1982); Benner et al., Modern metabolism as a palimpsest of the RNA world. Proc. Natl. Acad. Sci. USA 86, 7054-7058 (1989)) that TPP, coenzyme B12 and FMN emerged as biological cofactors during the RNA world (Joyce, The antiquity of RNA-based evolution. Nature 418, 214-221 (2002)). If these metabolites were being biosynthesized and used before the advent of proteins, then certain riboswitches might be modern examples of the most ancient form of genetic control. A search of genomic sequence databases has revealed that sequences corresponding to the TPP aptamer exist in organisms from bacteria, archaea and eukarya—largely without major alteration. Although new metabolite-binding mRNAs are likely to emerge as evolution progresses, it is possible that the known riboswitches are molecular fossils from the RNA world.

[0091] Disclosed are mRNA elements that have been identified in fungi and in plants that match the consensus sequence and structure of thiamine pyrophosphate-binding domains of prokaryotes. In Arabidopsis, the consensus motif resides in the 3’-UTR of a thiamine biosynthetic gene, and the isolated RNA domain binds the corresponding coenzyme in vitro.
These results indicate that metabolite-binding mRNAs are involved in eukaryotic gene regulation and that some riboswitches might be representatives of an ancient form of genetic control.

[0092] It is to be understood that the disclosed method and compositions are not limited to specific synthetic methods, specific analytical techniques, or to particular reagents unless otherwise specified, and, as such, can vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting.

Materials

[0093] Disclosed are materials, compositions, and components that can be used for, can be used in conjunction with, can be used in preparation for, or are products of the disclosed methods and compositions. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference to each of various individual and collective combinations and permutation of these compounds can not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a riboswitch or aptamer domain is disclosed and discussed and a number of modifications that can be made to a number of molecules and the riboswitch or aptamer domain are discussed, each and every combination and permutation of riboswitch or aptamer and the modifications that are possible are specifically contemplated unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited, each is individually and collectively contemplated. Thus, in this example, each of the combinations A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are specifically contemplated and should be considered disclosed from disclosure of A, B, and C; D, E, and F; and the example combination A-D. Likewise, any subset or combination of these is also specifically contemplated and disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E are specifically contemplated and should be considered disclosed from disclosure of A, B, and C; D, E, and F; and the example combination A-D. This concept applies to all aspects of the application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods, and that each such combination is specifically contemplated and should be considered disclosed.

A. Riboswitches

[0094] Riboswitches are expression control elements that are part of the RNA molecule to be expressed and that change state when bound by a trigger molecule. Riboswitches typically can be dissected into two separate domains: one that selectively binds the target (aptamer domain) and another that influences genetic control (expression platform domain). It is the dynamic interplay between these two domains that results in metabolite-dependent allosteric control of gene expression. Disclosed are isolated and recombinant riboswitches, recombinant constructs containing such riboswitches, heterologous sequences operably linked to such riboswitches, and cells and transgenic organisms harboring such riboswitches, riboswitch recombinant constructs, and riboswitches operably linked to heterologous sequences. The heterologous sequences can be, for example, sequences encoding proteins or peptides of interest, including reporter proteins or peptides. Preferred riboswitches are, or are derived from, naturally occurring riboswitches.

[0095] The disclosed riboswitches, including the derivatives and recombinant forms thereof, generally can be from any source, including naturally occurring riboswitches and riboswitches designed de novo. Any such riboswitches can be used in or with the disclosed methods. However, different types of riboswitches can be defined and some such sub-types can be useful in or with particular methods (generally as described elsewhere herein). Types of riboswitches include, for example, naturally occurring riboswitches, derivatives and modified forms of naturally occurring riboswitches, chimeric riboswitches, and recombinant riboswitches. A naturally occurring riboswitch is a riboswitch having the sequence of a riboswitch as found in nature. Such a naturally occurring riboswitch can be an isolated or recombinant form of the naturally occurring riboswitch as it occurs in nature. That is, the riboswitch has the same primary structure but has been isolated or engineered in a new genetic or nucleic acid context. Chimeric riboswitches can be made up of, for example, part of a riboswitch of any or of a particular class or type of riboswitch and part of a different riboswitch of the same or of any different class or type of riboswitch; part of a riboswitch of any or of a particular class or type of riboswitch and any non-ribswitch sequence or component. Recombinant riboswitches are riboswitches that have been isolated or engineered in a new genetic or nucleic acid context.

[0096] Different classes of riboswitches refer to riboswitches that have the same or similar trigger molecules or riboswitches that have the same or similar overall structure (predicted, determined, or a combination). Riboswitches of the same class generally, but need not, have both the same or similar trigger molecules and the same or similar overall structure.

[0097] Also disclosed are chimeric riboswitches containing heterologous aptamer domains and expression platform domains. That is, chimeric riboswitches are made up an aptamer domain from one source and an expression platform domain from another source. The heterologous sources can be from, for example, different specific riboswitches, different types of riboswitches, or different classes of riboswitches. The heterologous aptamers can also come from non-ribo- switch aptamers. The heterologous expression platform domains can also come from non-ribswitch sources.

[0098] Riboswitches can be modified from other known, developed or naturally occurring riboswitches. For example, switch domain portions can be modified by changing one or more nucleotides while preserving the known or predicted secondary, tertiary, or both secondary and tertiary structure of the riboswitch. For example, both nucleotides in a base pair can be changed to nucleotides that can also base pair. Changes that allow retention of base pairing are referred to herein as base pair conservative changes.

[0099] Modified or derivative riboswitches can also be produced using in vitro selection and evolution techniques. In general, in vitro evolution techniques as applied to riboswitches involve producing a set of variant riboswitches where part(s) of the riboswitch sequence is varied while other
parts of the riboswitch are held constant. Activation, deactivation or blocking (or other functional or structural criteria) of the set of variant riboswitches can then be assessed and those variant riboswitches meeting the criteria of interest are selected for use or further rounds of evolution. Useful base riboswitches for generation of variants are the specific and consensus riboswitches disclosed herein. Consensus riboswitches can be used to inform which part(s) of a riboswitch to vary for in vitro selection and evolution.

[0100] Also disclosed are modified riboswitches with altered regulation. The regulation of a riboswitch can be altered by operably linking an aptamer domain to the expression platform domain of the riboswitch (which is a chimeric riboswitch). The aptamer domain can then mediate regulation of the riboswitch through the action of, for example, a trigger molecule for the aptamer domain. Aptamer domains can be operably linked to expression platform domains of riboswitches in any suitable manner, including, for example, by replacing the normal or natural aptamer domain of the riboswitch with a new aptamer domain. Generally, any compound or condition that can activate, deactivate or block the riboswitch from which the aptamer domain is derived can be used to activate, deactivate or block the chimeric riboswitch.

[0101] Also disclosed are inactivated riboswitches. Riboswitches can be inactivated by covalently altering the riboswitch (by, for example, crosslinking parts of the riboswitch or coupling a compound to the riboswitch). Inactivation of a riboswitch in this manner can result from, for example, an alteration that prevents the trigger molecule for the riboswitch from binding, that prevents the change in state of the riboswitch upon binding of the trigger molecule, or that prevents the expression platform domain of the riboswitch from affecting expression upon binding of the trigger molecule.

[0102] Also disclosed are biosensor riboswitches. Biosensor riboswitches are engineered riboswitches that produce a detectable signal in the presence of their cognate trigger molecule. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Biosensor riboswitches can be designed for use in vivo or in vitro. For example, biosensor riboswitches operably linked to a reporter RNA that encodes a protein that serves as or is involved in producing a signal can be used in vivo by engineering a cell or organism to harbor a nucleic acid construct encoding the riboswitch/reporter RNA. An example of a biosensor riboswitch for use in vitro is a riboswitch that includes a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a biosensor riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch. Biosensor riboswitches can be used in various situations and platforms. For example, biosensor riboswitches can be used with solid supports, such as plates, chips, strips and wells.

[0103] Also disclosed are modified or derivative riboswitches that recognize new trigger molecules. New riboswitches and/or new aptamers that recognize new trigger molecules can be selected for, designed or derived from known riboswitches. This can be accomplished by, for example, producing a set of aptamer variants in a riboswitch, assessing the activation of the variant riboswitches in the presence of a compound of interest, selecting variant riboswitches that were activated (or, for example, the riboswitches that were the most highly or the most selectively activated), and repeating these steps until a variant riboswitch of a desired activity, specificity, combination of activity and specificity, or other combination of properties results.

[0104] Particularly useful aptamer domains can form a stem structure referred to herein as the P1 stem structure (or simply P1). The P1 stems of a variety of riboswitches are shown in FIG. 11 (and in other figures). The hybridizing strands in the P1 stem structure are referred to as the aptamer strand (also referred to as the P1a strand) and the control strand (also referred to as the P1b strand). The control strand can form a stem structure with both the aptamer strand and a sequence in a linked expression platform that is referred to as the regulated strand (also referred to as the P1c strand). Thus, the control strand (P1b) can form alternative stem structures with the aptamer strand (P1a) and the regulated strand (P1c).

[0105] In general, any aptamer domain can be adapted for use with any expression platform domain by designing or adapting a regulated strand in the expression platform domain to be complementary to the control strand of the aptamer domain. Alternatively, the sequence of the aptamer and control strands of an aptamer domain can be adapted so that the control strand is complementary to a functionally significant sequence in an expression platform. For example, the control strand can be adapted to be complementary to the Shine-Dalgarno sequence of an RNA such that, upon formation of a stem structure between the control strand and the SD sequence, the SD sequence becomes inaccessible to ribosomes, thus reducing or preventing translation initiation. Note that the aptamer strand would have corresponding changes in sequence to allow formation of a P1 stem in the aptamer domain.

[0106] As another example, a transcription terminator can be added to an RNA molecule (most conveniently in an untranslated region of the RNA) where part of the sequence of the transcription terminator is complementary to the control strand of an aptamer domain (the sequence will be the regulated strand). This will allow the control sequence of the aptamer domain to form alternative stem structures with the aptamer strand and the regulated strand, thus either forming or disrupting a transcription terminator stem upon activation or deactivation of the riboswitch. Any other expression element can be brought under the control of a riboswitch by similar design of alternative stem structures.

[0107] For transcription terminators controlled by riboswitches, the speed of transcription and spacing of the riboswitch and expression platform elements can be important for proper control. Transcription speed can be adjusted by, for example, by including polymerase pausing elements (e.g., a series of uridine residues) to pause transcription and allow the riboswitch to form and sense trigger molecules. For example, with the FMN riboswitch, if FMN is bound to its aptamer domain, then the antiterminator sequence is sequenced and is unavailable for formation of an antiterminator structure (FIG. 12). However, if FMN is absent, the antiterminator can form once its nucleotides emerge from the polymerase. RNAP then breaks free of the pause site only to reach...
another U-stretch and pause again. The transcriptional terminator then forms only if the terminator nucleotides are not tied up by the antiterminator.

[0108] Disclosed are regulatable gene expression constructs comprising a nucleic acid molecule encoding an RNA comprising a riboswitch operably linked to a coding region, wherein the riboswitch regulates expression of the RNA, wherein the riboswitch and coding region are heterologous. The riboswitch can comprise an aptamer domain and an expression platform domain, wherein the aptamer domain and the expression platform domain are heterologous. The riboswitch can comprise an aptamer domain and an expression platform domain, wherein the aptamer domain comprises a P1 stem, wherein the P1 stem comprises an aptamer strand and a control strand, wherein the expression platform domain comprises a regulated strand, wherein the regulated strand, the control strand, or both have been designed to form a stem structure.

[0109] Disclosed are riboswitches, wherein the riboswitch is a non-natural derivative of a naturally-occurring riboswitch. The riboswitch can comprise an aptamer domain and an expression platform domain, wherein the aptamer domain and the expression platform domain are heterologous. The riboswitch can be derived from a naturally-occurring guanine-responsive riboswitch, adenine-responsive riboswitch, lysine-responsive riboswitch, thiamine pyrophosphate-responsive riboswitch, adenosylcobalamin-responsive riboswitch, flavin mononucleotide-responsive riboswitch, or a S-adenosylmethionine-responsive riboswitch. The riboswitch can be activated by a trigger molecule, wherein the riboswitch produces a signal when activated by the trigger molecule.

[0110] Numerous riboswitches and riboswitch constructs are described and referred to herein. It is specifically contemplated that any specific riboswitch or riboswitch construct or group of riboswitches or riboswitch constructs can be excluded from some aspects of the invention disclosed herein. For example, fusion of the xpt-phus riboswitch with a reporter gene could be excluded from a set of riboswitches fused to reporter genes.

[0111] 1. Aptamer Domains

[0112] Aptamers are nucleic acid segments and structures that can bind selectively to particular compounds and classes of compounds. Riboswitches have aptamer domains that, upon binding of a trigger molecule result in a change the state or structure of the riboswitch. In functional riboswitches, the state or structure of the expression platform domain linked to the aptamer domain changes when the trigger molecule binds to the aptamer domain. Aptamer domains of riboswitches can be derived from any source, including, for example, natural aptamer domains of riboswitches, artificial aptamers, engineered, selected, evolved or derived aptamers or aptamer domains. Aptamers in riboswitches generally have at least one portion that can interact, such as by forming a stem structure, with a portion of the linked aptamer domain. This stem structure will either form or be disrupted upon binding of the trigger molecule.

[0113] Consensus aptamer domains of a variety of natural riboswitches are shown in FIG. 11. These aptamer domains (including all of the direct variants embodied therein) can be used in riboswitches. The consensus sequences and structures indicate variations in sequence and structure. Aptamer domains that are within the indicated variations are referred to herein as direct variants. These aptamer domains can be modified to produce modified or variant aptamer domains. Conservative modifications include any change in base paired nucleotides such that the nucleotides in the pair remain complementary. Moderate modifications include changes in the length of stems or of loops (for which a length or length range is indicated) of less than or equal to 20% of the length range indicated. Loop and stem lengths are considered to be "indicated" where the consensus structure shows a stem or loop of a particular length or where a range of lengths is listed or depicted. Moderate modifications include changes in the length of stems or of loops (for which a length or length range is not indicated) of less than or equal to 40% of the length range indicated. Moderate modifications also include and functional variants of unspecified portions of the aptamer domain. Unspecified portions of the aptamer domains are indicated by solid lines in FIG. 11.

[0114] The P1 stem and its constituent strands can be modified in adapting aptamer domains for use with expression platforms and RNA molecules. Such modifications, which can be extensive, are referred to herein as P1 modifications. P1 modifications include changes in the sequence and/or length of the P1 stem of an aptamer domain.

[0115] The aptamer domains shown in FIG. 11 (including any direct variants) are particularly useful as initial sequences for producing derived aptamer domains via in vitro selection or in vitro evolution techniques.

[0116] Aptamer domains of the disclosed riboswitches can also be used for any other purpose, and in any other context, as aptamers. For example, aptamers can be used to control ribozymes, other molecular switches, and any RNA molecule where a change in structure can affect function of the RNA.

[0117] 2. Expression Platform Domains

[0118] Expression platform domains are a part of riboswitches that affect expression of the RNA molecule that contains the riboswitch. Expression platform domains generally have at least one portion that can interact, such as by forming a stem structure, with a portion of the linked aptamer domain. This stem structure will either form or be disrupted upon binding of the trigger molecule. The stem structure generally either is, or prevents formation of, an expression regulatory structure. An expression regulatory structure is a structure that allows, prevents, enhances or inhibits expression of an RNA molecule containing the structure. Examples include Shine-Dalgarno sequences, initiation codons, transcription terminators, and stability and processing signals.

B. Trigger Molecules

[0119] Trigger molecules are molecules and compounds that can activate a riboswitch. This includes the natural or normal trigger molecule for the riboswitch and other compounds that can activate the riboswitch. Natural or normal trigger molecules are the trigger molecule for a given riboswitch in nature or, in the case of some non-natural riboswitches, the trigger molecule for which the riboswitch was designed or with which the riboswitch was selected (as in, for example, in vitro selection or in vitro evolution techniques). Non-natural trigger molecules can be referred to as non-natural trigger molecules.
C. Compounds

[0120] Also disclosed are compounds, and compositions containing such compounds, that can activate, deactivate or block a riboswitch. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Compounds can be used to activate, deactivate or block a riboswitch. The trigger molecule for a riboswitch (as well as other activating compounds) can be used to activate a riboswitch. Compounds other than the trigger molecule generally can be used to deactivate or block a riboswitch. Riboswitches can also be deactivated by, for example, removing trigger molecules from the presence of the riboswitch. A riboswitch can be blocked by, for example, binding of an analog of the trigger molecule that does not activate the riboswitch.

[0121] Also disclosed are compounds for altering expression of an RNA molecule, or of a gene encoding an RNA molecule, where the RNA molecule includes a riboswitch. This can be accomplished by bringing a compound into contact with the RNA molecule. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Thus, subjecting an RNA molecule of interest that includes a riboswitch to conditions that activate, deactivate or block the riboswitch can be used to alter expression of the RNA. Expression can be altered as a result of, for example, termination of transcription or blocking of ribosome binding to the RNA. Binding of a trigger molecule can, depending on the nature of the riboswitch, reduce or prevent expression of the RNA molecule or promote or increase expression of the RNA molecule.

[0122] Also disclosed are compounds for regulating expression of an RNA molecule, or of a gene encoding an RNA molecule. Also disclosed are compounds for regulating expression of a naturally occurring gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. If the gene is essential for survival of a cell or organism that harbors it, activating, deactivating or blocking the riboswitch can in death, stasis or debilitation of the cell or organism.

[0123] Also disclosed are compounds for regulating expression of an isolated, engineered or recombinant gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. If the gene encodes a desired expression product, activating or deactivating the riboswitch can be used to induce expression of the gene and thus result in production of the expression product. If the gene encodes an inducer or repressor of gene expression or of another cellular process, activation, deactivation or blocking of the riboswitch can result in induction, repression, or de-repression of other, regulated genes or cellular processes. Many such secondary regulatory effects are known and can be adapted for use with riboswitches. An advantage of riboswitches as the primary control for such regulation is that riboswitch trigger molecules can be small, non-antigenic molecules.

[0124] Also disclosed are methods of identifying compounds that activate, deactivate or block a riboswitch. For example, compounds that activate a riboswitch can be identified by bringing into contact a test compound and a riboswitch and assessing activation of the riboswitch. If the riboswitch is activated, the test compound is identified as a compound that activates the riboswitch. Activation of a riboswitch can be assessed in any suitable manner. For example, the riboswitch can be linked to a reporter RNA and expression, expression level, or change in expression level of the reporter RNA can be measured in the presence and absence of the test compound. As another example, the riboswitch can include a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch. As can be seen, assessment of activation of a riboswitch can be performed with the use of a control assay or measurement or without the use of a control assay or measurement. Methods for identifying compounds that deactivate a riboswitch can be performed in analogous ways.

[0125] Identification of compounds that block a riboswitch can be accomplished in any suitable manner. For example, an assay can be performed for assessing activation or deactivation of a riboswitch in the presence of a compound known to active or deactivate the riboswitch and in the presence of a test compound. If activation or deactivation is not observed as would be observed in the absence of the test compound, then the test compound is identified as a compound that blocks activation or deactivation of the riboswitch.

[0126] Also disclosed are compounds made by identifying a compound that activates, deactivates or blocks a riboswitch and manufacturing the identified compound. This can be accomplished by, for example, combining compound identification methods as disclosed elsewhere herein with methods for manufacturing the identified compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound.

[0127] Also disclosed are compounds made by checking activation, deactivation or blocking of a riboswitch by a compound and manufacturing the checked compound. This can be accomplished by, for example, combining compound activation, deactivation or blocking assessment methods as disclosed elsewhere herein with methods for manufacturing the checked compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound. Checking compounds for their ability to activate, deactivate or block a riboswitch refers to both identification of compounds previously unknown to activate, deactivate or block a riboswitch and to assessing the ability of a compound to activate, deactivate or block a riboswitch where the compound was already known to activate, deactivate or block the riboswitch.

[0128] Specific compounds that can be used to activate riboswitches are also disclosed. Compounds useful with guanine-responsive riboswitches (and riboswitches derived from guanine-responsive riboswitches) include compounds having the formula
where the compound can bind a guanine-responsive riboswitch or derivative thereof, where, when the compound is bound to a guanine-responsive riboswitch or derivative, \( R_7 \) serves as a hydrogen bond acceptor, \( R_{10} \) serves as a hydrogen bond donor, \( R_{11} \) serves as a hydrogen bond acceptor, \( R_{12} \) serves as a hydrogen bond donor, where \( R_{13} \) is \( H \), \( H \) or is not present, where \( R_1 \), \( R_2 \), \( R_3 \), \( R_4 \), \( R_5 \), \( R_6 \), \( R_9 \), and \( R_8 \) are each independently \( C \), \( N \), \( O \), or \( S \), and where \( \cdots \) each independently represent a single or double bond.

Every compound within the above definition is intended to be and should be considered to be specifically disclosed herein. Further, every subgroup that can be identified within the above definition is intended to be and should be considered to be specifically disclosed herein. As a result, it is specifically contemplated that any compound, or subgroup of compounds can be either specifically included for or excluded from use or included in or excluded from a list of compounds. For example, as one option, a group of compounds is contemplated where each compound is as defined above but is not guanine, hypoxanthine, xanthine, or \( N^+ \)-methylguanine. As another example, a group of compounds is contemplated where each compound is as defined above and is able to activate a guanine-responsive riboswitch.

Compounds useful with adenine-responsive riboswitches (and riboswitches derived from adenine-responsive riboswitches) include compounds having the formula

where the compound can bind a lysine-responsive riboswitch or derivative thereof, where \( R_2 \) and \( R_3 \) are each positively charged, where \( R_1 \) is negatively charged, where \( R_4 \) is \( C \), \( N \), \( O \), or \( S \), and where \( \cdots \) each independently represent a single or double bond. Also contemplated are compounds as defined above where \( R_2 \) and \( R_3 \) are each \( \text{NH}^+ \) and where \( R_1 \) is \( O^- \).

Every compound within the above definition is intended to be and should be considered to be specifically disclosed herein. Further, every subgroup that can be identified within the above definition is intended to be and should be considered to be specifically disclosed herein. As a result, it is specifically contemplated that any compound, or subgroup of compounds can be either specifically included for or excluded from use or included in or excluded from a list of compounds. For example, as one option, a group of compounds is contemplated where each compound is as defined above but is not lysine. As another example, a group of compounds is contemplated where each compound is as defined above and is able to activate a lysine-responsive riboswitch.

Compounds useful with TPP-responsive riboswitches (and riboswitches derived from lysine-responsive riboswitches) include compounds having the formula

where the compound can bind a TPP-responsive riboswitch or derivative thereof, where \( R_1 \) is positively charged, where \( R_2 \) and \( R_3 \) are each independently \( C \), \( O \), or \( S \), where \( R_4 \) is \( \text{CH}_2\text{NH}_2 \), \( \text{NH}_2 \), \( \text{OH} \), \( \text{SH} \) or \( H \) or not present, where \( R_5 \) is \( \text{CH}_2 \), \( \text{NH}_2 \), \( \text{OH} \), \( \text{SH} \) or \( H \), where \( R_6 \) is \( \text{C} \) or \( \text{N} \) and where \( \cdots \) each independently represent a single or double bond. Also contemplated are compounds as defined above where \( R_1 \) is phosphate, diphosphate or triphosphate.

Every compound within the above definition is intended to be and should be considered to be specifically disclosed herein. Further, every subgroup that can be identified within the above definition is intended to be and should be considered to be specifically disclosed herein. As a result, it is specifically contemplated that any compound, or subgroup of compounds can be either specifically included for or excluded from use or included in or excluded from a list of compounds. For example, as one option, a group of compounds is contemplated where each compound is as defined above but is not TPP, TP or thiamine. As another example, a
group of compounds is contemplated where each compound is as defined above and is able to activate a TPP-responsive riboswitch.

D. Constructs, Vectors and Expression Systems

[0138] The disclosed riboswitches can be used in any suitable expression system. Recombinant expression is usually accomplished using a vector, such as a plasmid. The vector can include a promoter operably linked to riboswitch-encoding sequence and RNA to be expression (e.g., RNA encoding a protein). The vector can also include other elements required for transcription and translation. As used herein, vector refers to any carrier containing exogenous DNA. Thus, vectors are agents that transport the exogenous nucleic acid into a cell without degradation and include a promoter yielding expression of the nucleic acid in the cells into which it is delivered. Vectors include but are not limited to plasmids, viral nucleic acids, viruses, plasmid nucleic acids, plages, cosmids, and artifical chromosomes. A variety of prokaryotic and eukaryotic expression vectors suitable for carrying riboswitch-regulated constructs can be produced. Such expression vectors include, for example, PET, PET3d, pCR2.1, pBAD, pUC, and yeast vectors. The vectors can be used, for example, in a variety of in vivo and in vitro situations.

[0139] Viral vectors include adenovirus, adenov-assoociated virus, herpes virus, vaccinia virus, polio virus, AIDS virus, neuronal trophic virus, Sindbis and other RNA viruses, including these viruses with the HIV backbone. Also useful are any viral families which share the properties of these viruses which make them suitable for use as vectors. Retroviral vectors, which are described in Verma (1985), include Murine Maloney Leukemia virus, MMLV, and retroviruses that express the desirable properties of MMLV as a vector. Typically, viral vectors contain nonstructural early genes, structural late genes, an RNA polymerase III transcript, inverted terminal repeats necessary for replication and encapsidation, and promoters to control the transcription and replication of the viral genome. When engineered as vectors, viruses typically have one or more of the early genes removed and a gene or gene promoter cassette is inserted into the viral genome in place of the removed viral DNA.

[0140] A “promoter” is generally a sequence or sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A “promoter” contains core elements required for basic interaction of RNA polymerase and transcription factors and can contain upstream elements and response elements.

[0141] “Enhancer” generally refers to a sequence of DNA that functions at no fixed distance from the transcription start site and can be either 5’ (Laimins, 1981) or 3’ (Lusky et al., 1983) to the transcription unit. Furthermore, enhancers can be within an intron (Banerji et al., 1983) as well as within the coding sequence itself (Osborne et al., 1984). They are usually between 10 and 300 bp in length, and they function in cis. Enhancers function to increase transcription from nearby promoters. Enhancers, like promoters, also often contain response elements that mediate the regulation of transcription. Enhancers often determine the regulation of expression.

[0142] Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human or nucleated cells) can also contain sequences necessary for the termination of transcription which can affect mRNA expression. These regions are transcribed as polyadenylated segments in the untranslated portion of the mRNA encoding tissue factor protein. The 3’ untranslated regions also include transcription termination sites. It is preferred that the transcription unit also contain a polyadenylation region. One benefit of this region is that it increases the likelihood that the transcribed unit will be processed and transported like mRNA. The identification and use of polyadenylation signals in expression constructs is well established. It is preferred that homologous polyadenylation signals be used in the transgene constructs.

[0143] The vector can include nucleic acid sequence encoding a marker product. This marker product is used to determine if the gene has been delivered to the cell and once delivered is being expressed. Preferred marker genes are the E. Coli lacZ gene which encodes β-galactosidase and green fluorescent protein.

[0144] In some embodiments the marker can be a selectable marker. When such selectable markers are successfully transferred into a host cell, the transformed host cell can survive if placed under selective pressure. There are two widely used distinct categories of selective regimes. The first category is based on a cell’s metabolism and the use of a mutant cell line which lacks the ability to grow independent of a supplemented media. The second category is dominant selection which refers to a selection scheme used in any cell type and does not require the use of a mutant cell line. These schemes typically use a drug to arrest growth of a host cell. Those cells which have a novel gene would express a protein conveying drug resistance and would survive the selection. Examples of such dominant selection use the drugs neomycin, (Southern and Berg, 1982), mycophenolic acid, (Mulligan and Berg, 1980) or hygromycin (Studgen et al., 1985).

[0145] Gene transfer can be obtained using direct transfer of genetic material, in but not limited to, plasmids, viral vectors, viral nucleic acids, phage nucleic acids, plages, cosmids, and artificial chromosomes, or via transfer of genetic material in cells or carriers such as cationic liposomes. Such methods are well known in the art and readily adaptable for use in the method described herein. Transfer vectors can be any nucleotide construction used to deliver genes into cells (e.g., a plasmid), or as part of a general strategy to deliver genes, e.g., as part of recombinant retroviruses or adenovirus (Ram et al. Cancer Res. 53:83-88, (1993)). Appropriate means for transfection, including viral vectors, chemical transfectants, or physico-mechanical methods such as electroporation and direct injection are well known in the art. For example, Wolff, J. A., et al., Science, 247, 1465-1468, (1990); and Wolff, J. A. Nature, 352, 815-818, (1991).

[0146] 1. Viral Vectors

[0147] Preferred viral vectors are Adenovirus, Adeno-associated virus, Herpes virus, Vaccinia virus, Polio virus, AIDS virus, neuronal trophic virus, Sindbis and other RNA viruses, including these viruses with the HIV backbone. Also preferred are any viral families which share the properties of these viruses which make them suitable for use as vectors. Preferred retroviruses include Murine Maloney Leukemia virus, MMLV, and retroviruses that express the desirable properties of MMLV as a vector. Retroviral vectors are able to carry a larger genetic payload, i.e., a transgene or marker gene, than other viral vectors, and for this reason are a commonly used vector. However, they are not useful in nonproliferating cells. Adenovirus vectors are relatively stable and easy to work with, have high titers, and can be delivered in aerosol formulation, and can transfect non-dividing cells. Pox viral vectors are large and have several sites for inserting genes, they are thermostable and can be stored at room tem-
perature. A preferred embodiment is a viral vector which has been engineered so as to suppress the immune response of the host organism, elicited by the viral antigens. Preferred vectors of this type will carry coding regions for Interleukin 8 or 10.

[0148] Viral vectors have higher transduction (ability to introduce genes) abilities than do most chemical or physical methods to introduce genes into cells. Typically, viral vectors contain nonstructural early genes, structural late genes, an RNA polymerase III transcript, inverted terminal repeats necessary for replication and encapsidation, and promoters to control the transcription and replication of the viral genome. When engineered as vectors, viruses typically have one or more of the early genes removed and a gene or gene/promotor cassette is inserted into the viral genome in place of the removed viral DNA. Constructs of this type can carry up to about 8 kb of foreign genetic material. The necessary functions of the removed early genes are typically supplied by cell lines which have been engineered to express the gene products of the early genes in trans.

[0149] i. Retroviral Vectors


[0151] A retrovirus is essentially a package which has packed into it a nucleic acid cargo. The nucleic acid cargo carries with it a packaging signal, which ensures that the replicated daughter molecules will be efficiently packaged within the package coat. In addition to the package signal, there are a number of molecules which are needed in cis, for the replication, and packaging of the replicated virus. Typically a retroviral genome, contains the gag, pol, and env genes which are involved in the making of the protein coat. It is the gag, pol, and env genes which are typically replaced by the foreign DNA that it is to be transferred to the target cell. Retrovirus vectors typically contain a packaging signal for incorporation into the package coat, a sequence which signals the start of the gag transcription unit, elements necessary for reverse transcription, including a primer binding site to bind the tRNA primer of reverse transcription, terminal repeat sequences that guide the switch of RNA strands during DNA synthesis, a purine rich sequence 5' to the 3' LTR that serve as the priming site for the synthesis of the second strand of DNA synthesis, and specific sequences near the ends of the LTRs that enable the insertion of the DNA state of the retrovirus to insert into the host genome. The removal of the gag, pol, and env genes allows for about 8 kb of foreign sequence to be inserted into the viral genome, become reverse transcribed, and upon replication be packaged into a new retroviral particle. This amount of nucleic acid is sufficient for the delivery of a one to many genes depending on the size of each transcript. It is preferable to include either positive or negative selectable markers along with other genes in the insert.

[0152] Since the replication machinery and packaging proteins in most retroviral vectors have been removed (gag, pol, and env), the vectors are typically generated by placing them into a packaging cell line. A packaging cell line is a cell line which has been transfected or transformed with a retrovirus that contains the replication and packaging machinery, but lacks any packaging signal. When the vector carrying the DNA of choice is transfected into these cell lines, the vector containing the gene of interest is replicated and packaged into new retroviral particles, by the machinery provided in cis by the helper cell. The genomes for the machinery are not packaged because they lack the necessary signals.

[0153] ii. Adenoviral Vectors


[0155] A preferred viral vector is one based on an adenovirus which has had the E1 gene removed and these viruses are generated in a cell line such as the human 293 cell line. In another preferred embodiment both the E1 and E3 genes are removed from the adenovirus genome.

[0156] Another type of viral vector is based on an adeno-associated virus (AAV). This defective parvovirus is a preferred vector because it can infect many cell types and is nonpathogenic to humans. AAV type vectors can transport about 4 to 5 kb and wild type AAV is known to stably insert into chromosome 19. Vectors which contain this site specific integration property are preferred. An especially preferred embodiment of this type of vector is the P4.1 C virus produced by Avigen, San Francisco, Calif., which can contain the herpes simplex virus thymidine kinase gene, HSV-tk, and/or a marker gene, such as the gene encoding the green fluorescent protein, GFP.

[0157] The inserted genes in viral and retroviral usually contain promoters, and/or enhancers to help control the
expression of the desired gene product. A promoter is generally a sequence or sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and can contain upstream elements and response elements.

[0158] 2. Viral Promoters and Enhancers

Preferred promoters controlling transcription from vectors in mammalian host cells can be obtained from various sources, for example, the genomes of viruses such as: polyoma, Simian Virus 40 (SV40), adenovirus, retroviruses, hepatitis-B virus and most preferably cytomegalovirus, or from heterologous mammalian promoters, e.g. beta actin promoter. The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment which also contains the SV40 viral origin of replication (Fiers et al., Nature, 273: 113 (1978)). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a HindIII restriction fragment (Greenway, P. J. et al., Gene 18: 355-360 (1982)). Of course, promoters from the host cell or related species also are useful herein.

[0160] Enhancer generally refers to a sequence of DNA that functions at no fixed distance from the transcription start site and can be either 5' (Laimins, L. et al., Proc. Natl. Acad. Sci. 78: 993 (1981)) or 3' (Linsky, M. I., et al., Mol. Cell. Bio. 3: 1108 (1983)) to the transcription unit. Furthermore, enhancers can be within an intron (Banerji, J. L. et al., Cell 33: 729 (1983)) as well as within the coding sequence itself (Osborne, T. F., et al., Mol. Cell. Bio. 4: 1293 (1984)). They are usually between 10 and 300 bp in length, and they function in cis. Enhancers function to increase transcription from nearby promoters. Enhancers also often contain response elements that mediate the regulation of transcription. Promoters can also contain response elements that mediate the regulation of transcription. Enhancers often determine the regulation of expression of a gene. While many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, alpha-fetoprotein and insulin), typically one will use an enhancer from a eukaryotic cell virus. Preferred examples are the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the poloma enhancer on the late side of the replication origin, and adenovirus enhancers.

[0161] The promoter and/or enhancer can be specifically activated either by light or specific chemical events which trigger their function. Systems can be regulated by reagents such as tetracycline and dexamethasone. There are also ways to enhance viral vector gene expression by exposure to irradiation, such as gamma irradiation, or alkylating chemotherapy drugs.

[0162] It is preferred that the promoter and/or enhancer region be active in all eukaryotic cell types. A preferred promoter of this type is the CMV promoter (650 bases). Other preferred promoters are SV40 promoters, cytomegalovirus (full length promoter), and retroviral vector LTR.

[0163] It has been shown that all specific regulatory elements can be cloned and used to construct expression vectors that are selectively expressed in specific cell types such as melanoma cells. The glial fibrillary acetic protein (GFAP) promoter has been used to selectively express genes in cells of glial origin.

[0164] Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human or nucleated cells) can also contain sequences necessary for the termination of transcription which can affect mRNA expression. These regions are transcribed as polyadenylated segments in the untranslated portion of the mRNA encoding tissue factor protein. The 3' untranslated regions also include transcription termination sites. It is preferred that the transcription unit also contain a polyadenylation region. One benefit of this region is that it increases the likelihood that the transcribed unit will be processed and transported like mRNA. The identification and use of polyadenylation signals in expression constructs is well established. It is preferred that homologous polyadenylation signals be used in the transgene constructs. In a preferred embodiment of the transcription unit, the polyadenylation region is derived from the SV40 early polyadenylation signal and consists of about 400 bases. It is also preferred that the transcribed units contain other standard sequences alone or in combination with the above sequences improve expression from, or stability of, the construct.

[0165] 3. Markers

[0166] The vectors can include nucleic acid sequence encoding a marker product. This marker product is used to determine if the gene has been delivered to the cell and once delivered is being expressed. Preferred marker genes are the E. Coli lacZ gene which encodes beta-galactosidase and green fluorescent protein.

[0167] In some embodiments the marker can be a selectable marker. Examples of suitable selectable markers for mammalian cells are dihydrofolate reductase (DHFR), thymidine kinase, neomycin, neomycin analog G418, hydromycin, and puromycin. When such selectable markers are successfully transferred into a mammalian host cell, the transformed mammalian host cell can survive if placed under selective pressure. There are two widely used distinct categories of selectable regimes. The first category is based on a cell's metabolism and the use of a mutant cell line which lacks the ability to grow independent of a supplemented media. Two examples are: CHO DHFR cells and mouse LTK cells. These cells lack the ability to grow without the addition of such nutrients as thymidine or hypoxanthine. Because these cells lack certain genes necessary for a complete nucleic acid synthesis pathway they cannot survive unless the missing nucleotides are provided in a supplemented media. An alternative to supplementing the media is to introduce an intact DHFR or TK gene into cells lacking the respective genes, thus altering their growth requirements. Individual cells which were not transformed with the DHFR or TK gene will not be capable of survival in non-supplemented media.

[0168] The second category is dominant selection which refers to a selection scheme used in any cell type and does not require the use of a mutant cell line. These schemes typically use a drug to arrest growth of a host cell. Those cells which would express a protein conveying drug resistance and would survive the selection. Examples of such dominant selection are the drugs neomycin, (Southern P. and Berg, P., J. Molec. Appl. Genet. 1: 327 (1982)), mycophenolic acid, (Mulligan, R. C. and Berg, P. Science 209: 1422 (1980)) or hygromycin, (Sugden, B. et al., Mol. Cell. Biol. 5: 410-413 (1985)). The three examples employ bacterial genes under eukaryotic control to convey resistance to the appropriate drug G418 or neomycin (geneticin), xgp7 (mycophenolic acid) or hygromycin, respectively. Others include the neomycin analog G418 and puromycin.

E. Biosensor Riboswitches

[0169] Also disclosed are biosensor riboswitches. Biosensor riboswitches are engineered riboswitches that produce a
detectable signal in the presence of their cognate trigger molecule. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Biosensor riboswitches can be designed for use in vivo or in vitro. For example, biosensor riboswitches operably linked to a reporter RNA that encodes a protein that serves as, or is involved in producing a signal can be used in vivo by engineering a cell or organism to harbor a nucleic acid construct encoding the riboswitch/reporter RNA. An example of a biosensor riboswitch for use in vitro is a riboswitch that includes a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a biosensor riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch.

F. Reporter Proteins and Peptides

[0170] For assessing activation of a riboswitch, or for biosensor riboswitches, a reporter protein or peptide can be used. The reporter protein or peptide can be encoded by the DNA the expression of which is regulated by the riboswitch. The examples describe the use of some specific reporter proteins. The use of reporter proteins and peptides is well known and can be adapted easily for use with riboswitches. The reporter proteins can be any protein or peptide that can be detected or that produces a detectable signal. Preferably, the presence of the protein or peptide can be detected using standard techniques (e.g., radioimmunoassay, radio-labeling, immunoassay, assay for enzymatic activity, absorbance, fluorescence, luminescence, and Western blot). More preferably, the level of the reporter protein is easily quantifiable using standard techniques even at low levels. Useful reporter proteins include luciferases, green fluorescent proteins and their derivatives, such as firefly luciferase (FL) from Photinus pyralis, and Renilla luciferase (RL) from Renilla reniformis.

G. Conformation Dependent Labels

[0171] Conformation dependent labels refer to all labels that produce a change in fluorescence intensity or wavelength based on a change in the form or conformation of the molecule or compound (such as a riboswitch) with which the label is associated. Examples of conformation dependent labels used in the context of probes and primers include molecular beacons, Amplifluors, FRET probes, cleavable FRET probes, TaqMan probes, scorpion primers, fluorescent triplex oligos including but not limited to triplex molecular beacons or triplex FRET probes, fluorescent water-soluble conjugated polymers, PNA probes and QPNA probes. Such labels, and, in particular, the principles of their function, can be adapted for use with riboswitches. Several types of conformation dependent labels are reviewed in Schweitzer and Kingsmore, Curr. Opin. Biotech. 12:21-27 (2001).

[0172] Stem quenched labels, a form of conformation dependent labels, are fluorescent labels positioned on a nucleic acid such that when a stem structure forms a quenching moiety is brought into proximity such that fluorescence from the label is quenched. When the stem is disrupted (such as when a riboswitch containing the label is activated), the quenching moiety is no longer in proximity to the fluorescent label and fluorescence increases. Examples of this effect can be found in molecular beacons, fluorescent triplex oligos, triplex molecular beacons, triplex FRET probes, and QPNA probes, the operational principles of which can be adapted for use with riboswitches.

[0173] Stem activated labels, a form of conformation dependent labels, are labels or pairs of labels where fluorescence is increased or altered by formation of a stem structure. Stem activated labels can include an acceptor fluorescent label and a donor moiety such that, when the acceptor and donor are in proximity (when the nucleic acid strands containing the labels form a stem structure), fluorescence resonance energy transfer from the donor to the acceptor causes the acceptor to fluoresce. Stem activated labels are typically pairs of labels positioned on nucleic acid molecules (such as riboswitches) such that the acceptor and donor are brought into proximity when a stem structure is formed in the nucleic acid molecule. If the donor moiety of a stem activated label is itself a fluorescent label, it can release energy as fluorescence (typically at a different wavelength than the fluorescence of the acceptor) when not in proximity to an acceptor (that is, when a stem structure is not formed). When the stem structure forms, the overall effect would then be a reduction of donor fluorescence and an increase in acceptor fluorescence. FRET probes are an example of the use of stem activated labels, the operational principles of which can be adapted for use with riboswitches.

H. Detection Labels

[0174] To aid in detection and quantitation of riboswitch activation, deactivation or blocking, or expression of nucleic acids or protein produced upon activation, deactivation or blocking of riboswitches, detection labels can be incorporated into detection probes or detection molecules or directly incorporated into expressed nucleic acids or proteins. As used herein, a detection label is any molecule that can be associated with nucleic acid or protein, directly or indirectly, and which results in a measurable, detectable signal, either directly or indirectly. Many such labels are known to those of skill in the art. Examples of detection labels suitable for use in the disclosed method are radioactive isotopes, fluorescent molecules, phosphorescent molecules, enzymes, antibodies, and ligands.

[0175] Examples of suitable fluorescent labels include fluorescein isothiocyanate (FITC), 5,6-carboxyfluorescein, Texas red, nitrobenz-2-oxa-1,3-diazol-4-yl (NBD), coumarin, dansyl chloride, rhodamine, amino-methyl coumarin (AMCA), Eosin, Erythrosin, BODIPY®, Cascade Blue®, Oregon Green®, pyrene, lissamine, xanthene, acidine, oxazines, phycerythrin, macrocyclic chelates of lanthanide ions such as quantum Dye™, fluorescent energy transfer dyes, such as thiazole orange-ethidium heterodimer, and the cyanine dyes Cy3, Cy5.5, Cy5, Cy5.5 and Cy7. Examples of other specific fluorescent labels include 3-Hydroxypropene 5,8,10-Tri Sulfonic acid, 5-Hydroxy Tryptamine (5-HT), Acid Fuchsin, Alizarin Complexon, Alizarin Red, Allophycocyanin, Aminoacoumarin, Anthroyl Stearate, AstraZen Brilliant Red 4G, AstraZen Orange R, AstraZen Red 6B, AstraZen Yellow 7 GLL, Atbrune, Auramine, Aurophosphine, Aurophosphine G, BAO 9 (Bisaminophenoxadiazole), BCECF, Berberine Sulphate, Bisbenzamide, Blanchophor FFG Solution, Blanchophor SV, Bodipy F1, Brilliant Sulphophlavin FF, Calcein Blue, Calcium Green, Calceofhor RW Solution, Calceofhor White, Calceophor White ABT Solution, Calceophor White Standard Solution, Carboxostyril, Cascade Yellow, Catecholamine, Chinacrine, Coriphosphine O, Coumarin-Phalloidin, CY3.1,8, CY5.1,8, CY7, Dans (1-Dimethyl Amino Naphthaline 5 Sulphonic Acid),Dansa (Diamino Naphyl Sulphonic Acid),

[0176] Useful fluorescent labels are fluorescein (5-carboxyfluorescein-N-hydroxysuccinimide ester), rhodamine (5,6-tetramethyl rhodamine), and the cyanine dyes Cy3, Cy4, Cy5, Cy5.5, and Cy7. The absorption and emission maxima, respectively, for these dyes are: FITC (490 nm; 520 nm), Cy3 (554 nm; 568 nm), Cy5.5 (581 nm; 588 nm), Cy5 (652 nm; 672 nm), Cy5 (682 nm; 703 nm) and Cy7 (755 nm; 778 nm), thus allowing their simultaneous detection. Other examples of fluorescent dyes include 6-carboxyfluorescein (6-FAM), 2',4',1,4-tetraethylrhodamine (TET), 2',4',5',7'-1,4-hexachlorofluorescein (HEX), 2',7'-dimethoxy-4,5'-dichloro-6-carboxyfluorescein (JOE), 2',4',5'-fluoro-7',8'-fused phen-1,4-dichloro-6-carboxyfluorescein (NED), and 2',4',5'-fluoro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein (VIC). Fluorescein labels can be obtained from a variety of commercial sources, including Amersham Pharmacia Biotech, Piscataway, N.J.; Molecular Probes, Eugene, Ore.; and Research Organics, Cleveland, Ohio.

[0177] Additional labels of interest include those that provide for signal only when the probe with which they are associated is specifically bound to a target molecule, where such labels include: “molecular beacons” as described in Tyagi & Kramer, Nature Biotechnology (1996) 14:303 and EP 0 070 685 B1. Other labels of interest include those described in U.S. Pat. No. 5,563,037; WO 97/17471 and WO 97/17076.

[0178] Labeled nucleotides are a useful form of detection label for direct incorporation into expressed nucleic acids during synthesis. Examples of detection labels that can be incorporated into nucleic acids include nucleotide analogs such as BrdUrd (5-bromo-2'-deoxyuridine), Hyb and Schmitke, Mutation Research 290:217-230 (1993)), unaminoallyldeoxyuridine (Henegarri et al., Nature Biotechnology 18:345-348 (2000)), 5-methylcytosine (Sano et al., Biochim. Biophys. Acta 951:157-165 (1988)), bromouridine (Wansick et al., J. Cell Biology 122:283-293 (1993)) and nucleotides modified with biotin (Langer et al., Proc. Natl. Acad. Sci. USA 78:6633 (1981)) or with suitable haptons such as digoxigenin (Kerkhof, Anal. Biochem. 205:359-364 (1992)). Suitable fluorescence-labeled nucleotides are Fluorescein-iso-thiocyanate-dUTP, Cyanine-3-dUTP and Cyamine-5-dUTP (Yu et al., Nucleic Acids Res. 22:3226-3232 (1994)). A preferred nucleotide analog detection label for DNA is BrdUrd (bromo-2'-deoxyuridine, BrdUrd, BrdU, BUDr, Sigma-Aldrich Co.). Other useful nucleotide analogs for incorporation of detection label into DNA are AA-dUTP (aminopyrrole-2-deoxyuridine triphosphate, Sigma-Aldrich Co.), and 5-methyl-dCTP (Roche Molecular Biochemicals). A useful nucleotide analog for incorporation of detection label into RNA is biotin-16-UTP (biotin-16-uridine-5-triphosphate, Roche Molecular Biochemicals). Fluorescein, Cy3, and Cy5 can be linked to dUTP for direct labelling. Cy3.5 and Cy7 are available as avidin or anti-digoxigenin conjugates for secondary detection of biotin- or digoxigenin-labelled probes.

[0179] Detection labels that are incorporated into nucleic acid, such as biotin, can be subsequently detected using sensitive methods well-known in the art. For example, biotin can be detected using streptavidin-alkaline phosphatase conjugate (Tropix, Inc.), which is bound to the biotin and subsequently detected by chemiluminescence of suitable substrates for chemiluminescence substrates (for example, chemiluminescence substrate CSPD: disodium 3,4-(methoxyphosphoryl)[1,2-5,6-dioxanone-3,2'-[5-chloroterephthalic][3,3,1.1.3]-decane-4-yl) phenyl phosphate; Tropix, Inc.). Labels can also be enzymes, such as alkaline phosphatase, soybean peroxidase, horseradish peroxidase and polymerases, that can be detected, for example, with chemical signal amplification or by using a substrate to the enzyme which produces light (for example, a chemiluminescence 1,2-dioxetane substrate) or fluorescent signal.

[0180] Molecules that combine two or more of these detection labels are also considered detection labels. Any of the known detection labels can be used with the disclosed probes, tags, molecules and methods to label and detect activated or deactivated riboswitches or nucleic acid or protein produced in the disclosed methods. Methods for detecting and measuring signals generated by detection labels are also known to those of skill in the art. For example, radioactive isotopes can be detected by scintillation counting or direct visualization; fluorescent molecules can be detected with fluorescent spectrophotometers; phosphorescent molecules can be detected with a spectrophotometer or directly visualized with a camera; enzymes can be detected by detection or visualization of the product of a reaction catalyzed by the enzyme; antibodies can be detected by detecting a secondary detection label coupled to the antibody. As used herein, detection molecules are molecules which interact with a compound or composition to be detected and to which one or more detection labels are coupled.

I. Sequence Similarities

[0181] It is understood that as discussed herein the use of the terms homology and identity mean the same thing as similarity. Thus, for example, if the use of the word homology is used between two sequences (non-natural sequences, for example) it is understood that this is not necessarily indicating an evolutionary relationship between these two sequences, but rather is looking at the similarity or related-
ness between their nucleic acid sequences. Many of the methods for determining homology between two evolutionarily related molecules are routinely applied to any two or more nucleic acids or proteins for the purpose of measuring sequence similarity regardless of whether they are evolutionarily related or not.

[0182] In general, it is understood that one way to define any known variants and derivatives or those that might arise, of the disclosed riboswitches, aptamers, expression platforms, genes and proteins herein, is through defining the variants and derivatives in terms of homology to specific known sequences. This identity of particular sequences disclosed herein is also discussed elsewhere herein. In general, variants of riboswitches, aptamers, expression platforms, genes and proteins herein disclosed typically have at least, about 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent homology to a stated sequence or a native sequence. Those of skill in the art readily understand how to determine the homology of two proteins or nucleic acids, such as genes. For example, the homology can be calculated after aligning the two sequences so that the homology is at its highest level.


[0184] The same types of homology can be obtained for nucleic acids by for example the algorithms disclosed in Zuker, M. Science 244:48-52, 1989, Jaeger et al. Proc. Natl. Acad. Sci. U.S.A 86:7706-7710, 1990, Jaeger et al. Methods Enzymol. 183:281-306, 1990 which are herein incorporated by reference for at least material related to nucleic acid alignment. It is understood that any of the methods typically can be used and that in certain instances the results of these various methods can differ, but the skilled artisan understands if identity is found with at least one of these methods, the sequences would be said to have the stated identity.

[0185] For example, as used herein, a sequence recited as having a particular percent homology to another sequence refers to sequences that have the recited homology as calculated by any one or more of the calculation methods described above. For example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using the Zuker calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by any of the other calculation methods. As another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using both the Zuker calculation method and the Pearson and Lipman calculation method even if the first sequence does not have 80 percent homology to the second sequence as calculated by the Smith and Waterman calculation method, the Needleman and Wunsch calculation method, the Jaeger calculation methods, or any of the other calculation methods. As yet another example, a first sequence has 80 percent homology, as defined herein, to a second sequence if the first sequence is calculated to have 80 percent homology to the second sequence using each of calculation methods (although, in practice, the different calculation methods will often result in different calculated homology percentages).

J. Hybridization and Selective Hybridization

[0186] The term hybridization typically means a sequence driven interaction between at least two nucleic acid molecules, such as a primer or a probe and a riboswitch or a gene. Sequence driven interaction means an interaction that occurs between two nucleotides or nucleotide analogs or nucleotide derivatives in a nucleotide specific manner. For example, G interacting with C or A interacting with T are sequence driven interactions. Typically sequence driven interactions occur on the Watson-Crick face or Hoogsteen face of the nucleotide. The hybridization of two nucleic acids is affected by a number of conditions and parameters known to those of skill in the art. For example, the salt concentrations, pH, and temperature of the reaction all affect whether two nucleic acid molecules will hybridize.

[0187] Parameters for selective hybridization between two nucleic acid molecules are well known to those of skill in the art. For example, in some embodiments selective hybridization conditions can be defined as stringent hybridization conditions. For example, stringency of hybridization is controlled by both temperature and salt concentration of either or both of the hybridization and washing steps. For example, the conditions of hybridization to achieve selective hybridization can involve hybridization in high ionic strength solution (6xSSC or 6xSSPE) at a temperature that is about 12-25°C below the Tm (the melting temperature at which half of the molecules dissociate from their hybridization partners) followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is about 5°C to 20°C below the Tm. The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations. The conditions can be used as described above to achieve stringency, or as is known in the art (Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989; Kunke et al. Methods Enzymol. 187:154:367, 1987 which is herein incorporated by reference for material at least related to hybridization of nucleic acids). A preferable stringent hybridization condition for a DNA:DNA hybridization can be at about 68°C. (in aqueous solution) in 6xSSC or 6xSSPE followed by washing at 68°C. Stringency of hybridization and washing, if desired, can be reduced accordingly as the degree of complementarity desired is decreased, and further, depending upon the G-C or A-T richness of any area wherein variability is searched for. Likewise, stringency of hybridization and washing, if desired, can be increased accordingly as homology desired is increased, and further, depending upon the G-C or A-T richness of any area wherein high homology is desired, all as known in the art.

[0188] Another way to define selective hybridization is by looking at the amount (percentage) of one of the nucleic acids
bound to the other nucleic acid. For example, in some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the limiting nucleic acid is bound to the non-limiting nucleic acid. Typically, the non-limiting nucleic acid is in for example, 10 fold or 100 fold or 1000 fold below their k_s, or where only one of the nucleic acid molecules is 10 fold or 100 fold or 1000 fold or where one or both nucleic acid molecules are above their k_s.

[0189] Another way to define selective hybridization is by looking at the percentage of nucleic acid that gets enzymatically manipulated under conditions where hybridization is required to promote the desired enzymatic manipulation. For example, in some embodiments selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the nucleic acid is enzymatically manipulated under conditions which promote the enzymatic manipulation, for example if the enzymatic manipulation is DNA extension, then selective hybridization conditions would be when at least about, 60, 65, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 percent of the nucleic acid molecules are extended. Preferred conditions also include those suggested by the manufacturer or indicated in the art as being appropriate for the enzyme performing the manipulation.

[0190] Just as with homology, it is understood that there are a variety of methods herein disclosed for determining the level of hybridization between two nucleic acid molecules. It is understood that these methods and conditions can provide different percentages of hybridization between two nucleic acid molecules, but unless otherwise indicated meeting the parameters of any of the methods would be sufficient. For example if 80% hybridization was required and as long as hybridization occurs within the required parameters in any one of these methods it is considered disclosed herein.

[0191] It is understood that those of skill in the art understand that if a composition or method meets any one of these criteria for determining hybridization either collectively or singly it is a composition or method that is disclosed herein.

K. Nucleic Acids

[0192] There are a variety of molecules disclosed herein that are nucleic acid based, including, for example, riboswitches, aptamers, and nucleic acids that encode riboswitches and aptamers. The disclosed nucleic acids can be made up of for example, nucleotides, nucleotide analogs, or nucleotide substitutes. Non-limiting examples of these and other molecules are discussed herein. It is understood that for example, when a vector is expressed in a cell, that the expressed mRNA will typically be made up of A, C, G, and U. Likewise, it is understood that if a nucleic acid molecule is introduced into a cell or cell environment through for example exogenous delivery, it is advantageous that the nucleic acid molecule be made up of nucleotide analogs that reduce the degradation of the nucleic acid molecule in the cellular environment.

[0193] So long as their relevant function is maintained, riboswitches, aptamers, expression platforms and any other oligonucleotides and nucleic acids can be made up of or include modified nucleotides (nucleotide analogs). Many modified nucleotides are known and can be used in oligonucleotides and nucleic acids. A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to the base moiety would include natural and synthetic modifications of A, C, G, and T/ U as well as different purine or pyrimidine bases, such as uracil-5-yl, hypoxanthin-9-yl (I), and 2-amino adenin-9-yl. A modified base includes but is not limited to 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-amino adenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiouracil and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azauracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxy and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazadenine and 3-deazaguanine and 3-deazadenine. Additional base modifications can be found for example in U.S. Pat. No. 3,687,808, Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613, and Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Langer, R., CRC Press, 1993. Certain nucleotide analogs, such as 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine can increase the stability of duplex formation. Other modified bases are those that function as universal bases. Universal bases include 3-nitropyrrole and 5-nitroindole. Universal bases substitute for the normal bases but have no bias in base pairing. That is, universal bases can base pair with any other base. Base modifications often can be combined with for example a sugar modification, such as 2'-O-methoxyethyl, to achieve unique properties such as increased duplex stability. There are numerous United States patents such as U.S. Pat. Nos. 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,666; 5,432,772; 5,487,187; 5,450,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121; 5,596,091; 5,614,617; and 5,681,941, which detail and describe a range of base modifications. Each of these patents is herein incorporated by reference in its entirety, and specifically for their description of base modifications, their synthesis, their use, and their incorporation into oligonucleotides and nucleic acids.

[0194] Nucleotide analogs can also include modifications of the sugar moiety. Modifications to the sugar moiety would include natural modifications of the ribose and deoxyribose as well as synthetic modifications. Sugar modifications include but are not limited to the following modifications at the 2' position: OH; F; O-S-, or N-alkyl; O-S-, or N-alkenyl; 0- S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl can be substituted or unsubstituted C1 to C10, alkyl or C2 to C10 alkenyl and alkynyl. 2' sugar modifications also include but are not limited to —O(CH2)n ONH(CH2)m, —O(CH2)n ON(CH2)m, —O(CH2)n ONH(CH2)m, —O(CH2)n ON(CH2)m, where n and m are from 1 to about 10.

[0195] Other modifications at the 2' position include but are not limited to: C1 to C10 lower alkyl, substituted lower alkyl,
alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH₂, OCN, Cl, Br, CN, CF₃, OCF₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, heterocycloalkyl, heterocycloalkaryl, aminocycloalkyl, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. Similar modifications can also be made at other positions on the sugar, particularly the 3\' position of the sugar on the 3\' terminal nucleotide or in 2\'-5\' linked oligonucleotides and the 5\' position of 5\' terminal nucleotide. Modified sugars would also include those that contain modifications at the bridging ring oxygen, such as CH₂ and S. Nucleotide sugar analogs can also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar. There are numerous United States patents that teach the preparation of such modified sugar structures such as U.S. Pat. Nos. 4,981,957; 5,118,800; 5,319,080; 5,389,044; 5,393,878; 5,446,137; 5,446,786; 5,514,785; 5,521,134; 5,367,811; 5,575,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; and 5,700,920, each of which is herein incorporated by reference in its entirety, and specifically for their description of modified sugar structures, their synthesis, their use, and their incorporation into nucleotides, oligonucleotides and nucleic acids.

[0196] Nucleotide analogs can also be modified at the phosphate moiety. Modified phosphate moieties include but are not limited to those that can be modified so that the linkage between two nucleotides contains a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphorothiester, aminalyphosphothiester, methyl and other alkyl phosphonates including 3\'-alkylphosphonate and chiral phosphonates, phosphonamidates including 3\'-amino phosphoramidate and aminolymphosphoramidates, thionophosphoramide, thionolymphosphoramide, and boranophosphates. It is understood that these phosphate or modified phosphate linkages between two nucleotides can be through a 3\'-5\' linkage or a 2\'-5\' linkage, and the linkage can contain inverted polarity such as 3\'-5\' to 5\'-3\' or 2\'-5\' to 5\'-2\'. Various salts, mixed salts and free acid forms are also included. Numerous United States patents teach how to make and use nucleotides containing modified phosphates and include but are not limited to, U.S. Pat. Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,288,717; 5,321,131; 5,389,767; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, each of which is herein incorporated by reference in its entirety, and specifically for their description of modified phosphates, their synthesis, their use, and their incorporation into nucleotides, oligonucleotides and nucleic acids.

[0197] It is understood that nucleotide analogs need only contain a single modification, but can also contain multiple modifications within one of the moieties or between different moieties.

[0198] Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize and hybridize to (base pair to) complementary nucleic acids in a Watson-Crick or Hoogsteen manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

[0199] Nucleotide substitutes are nucleotides or nucleotide analogs that have had the phosphate moiety and/or sugar moieties replaced. Nucleotide substitutes do not contain a standard phosphorus atom. Substitutes for the phosphate can be for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleinemino and methylenyldrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Numerous United States patents disclose how to make and use these types of phosphate replacements and include but are not limited to U.S. Pat. Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,633; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,480,677; 5,541,307; 5,561,225; 5,596,056; 5,602,240; 5,610,289; 5,622,240; 5,624,866; 5,633,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference in its entirety, and specifically for their description of phosphate replacements, their synthesis, their use, and their incorporation into nucleotides, oligonucleotides and nucleic acids.

[0200] It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced, by for example an amide type linkage (aminoethylglycine) (PNA). U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262 teach how to make and use PNA molecules, each of which is herein incorporated by reference. (See also Nielsen et al., Science 254:1497-1500 (1991)).

[0201] Oligonucleotides and nucleic acids can be comprised of nucleotides and can be made up of different types of nucleotides or the same type of nucleotides. For example, one or more of the nucleotides in an oligonucleotide can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; about 10% to about 50% of the nucleotides can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; about 50% or more of the nucleotides can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; or all of the nucleotides are ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides. Such oligonucleotides and nucleic acids can be referred to as chimeric oligonucleotides and chimeric nucleic acids.

L. Solid Supports

[0202] Solid supports are solid-state substrates or supports with which molecules (such as trigger molecules) and riboswitches (or other components used in, or produced by, the disclosed methods) can be associated. Riboswitches and other molecules can be associated with solid supports directly or indirectly. For example, analytes (e.g., trigger molecules, test compounds) can be bound to the surface of a solid support or associated with capture agents (e.g., compounds or mol-
ecules that bind an analyte) immobilized on solid supports. As another example, riboswitches can be bound to the surface of a solid support or associated with probes immobilized on solid supports. An array is a solid support to which multiple riboswitches, probes or other molecules have been associated in an array, grid, or other organized pattern.

Solid-state substrates for use in solid supports can include any solid material with which components can be associated, directly or indirectly. This includes materials such as acrylicamide, agarose, cellulose, nitrocellulose, glass, gold, polystyrene, polyethylene vinyl acetate, polypropylene, polyethylene glycol, polyethylene oxide, poly-silicates, polycarbonate, teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polyactic acid, polystyrene, functionalized silane, polypolyfluorurate, collagen, glycaminoglycans, and polyamino acids. Solid-state substrates can have any useful form including thin film, membrane, bottles, dishes, fibers, woven fibers, shaped polymers, particles, beads, microparticles, or a combination. Solid-state substrates and solid supports can be porous or non-porous. A chip is a rectangular or square small piece of material. Preferred forms for solid-state substrates are thin films, beads, or chips. A useful form for a solid-state substrate is a microtiter dish. Some embodiments, a multilayer glass slide can be employed.

An array can include a plurality of riboswitches, trigger molecules, other molecules, compounds or probes immobilized at identified or predefined locations on the solid support. Each predefined location on the solid support generally has one type of component (that is, all the components at that location are the same). Alternatively, multiple types of components can be immobilized in the same predefined location on a solid support. Each location will have multiple copies of the given components. The spatial separation of different components on the solid support allows separate detection and identification.

Although useful, it is not required that the solid support be a single unit or structure. A set of riboswitches, trigger molecules, other molecules, compounds and/or probes can be distributed over any number of solid supports. For example, at one extreme, each component can be immobilized in a separate reaction tube or container, or on separate beads or microparticles.


Each of the components (for example, riboswitches, trigger molecules, or other molecules) immobilized on the solid support can be located in a different predefined region of the solid support. The different locations can be different reaction chambers. Each of the different predefined regions can be physically separated from each other of the different regions. The distance between the different predefined regions of the solid support can be either fixed or variable. For example, in an array, each of the components can be arranged at fixed distances from each other, while components associated with beads will not be in a fixed spatial relationship. In particular, the use of multiple solid support units (for example, multiple beads) will result in variable distances.

Components can be associated or immobilized on a solid support at any density. Components can be immobilized to the solid support at a density exceeding 400 different components per cubic centimeter. Arrays of components can have an mumber of components. For example, an array can have at least 1,000 different components immobilized on the solid support, at least 10,000 different components immobilized on the solid support, at least 100,000 different components immobilized on the solid support, or at least 1,000,000 different components immobilized on the solid support.

M. Kits

The materials described above as well as other materials can be packaged together in any suitable combination as a kit useful for performing, or aiding in the performance of, the disclosed method. It is useful if the kit components in a given kit are designed and adapted for use together in the disclosed method. For example disclosed are kits for detecting compounds, the kit comprising one or more biosensor riboswitches. The kits also can contain reagents and labels for detecting activation of the riboswitches.

N. Mixtures

Disclosed are mixtures formed by performing or preparing to perform the disclosed method. For example, disclosed are mixtures comprising riboswitches and trigger molecules.

Whenever the method involves mixing or bringing into contact compositions or components or reagents, performing the method creates a number of different mixtures. For example, if the method includes 3 mixing steps, after each one of these steps a unique mixture is formed if the steps are performed separately. In addition, a mixture is formed at the completion of all of the steps regardless of how the steps were performed. The present disclosure contemplates these mixtures, obtained by the performance of the disclosed methods as well as mixtures containing any disclosed reagent, composition, or component, for example, disclosed herein.

O. Systems

Disclosed are systems useful for performing, or aiding in the performance of, the disclosed method. Systems generally comprise combinations of articles of manufacture such as structures, machines, devices, and the like, and compositions, compounds, materials, and the like. Such combinations that are disclosed or that are apparent from the disclosure are contemplated. For example, disclosed and contemplated are systems comprising sensor riboswitches, a solid support and a signal-reading device.

P. Data Structures and Computer Control

Disclosed are data structures used in, generated by, or generated from, the disclosed method. Data structures generally are any form of data, information, and/or objects collected, organized, stored, and/or embodied in a composition or medium. Riboswitch structures and activation measurements stored in electronic form, such as in RAM or on a storage disk, is a type of data structure.
[0214] The disclosed method, or any part thereof or preparation therefor, can be controlled, managed, or otherwise assisted by computer control. Such computer control can be accomplished by a computer controlled process or method, can use and/or generate data structures, and can use a computer program. Such computer control, computer controlled processes, data structures, and computer programs are contemplated and should be understood to be disclosed herein.

Methods

[0215] Disclosed are methods for activating, deactivating or blocking a riboswitch. Such methods can involve, for example, bringing into contact a riboswitch and a compound or trigger molecule that can activate, deactivate or block the riboswitch. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Compounds can be used to activate, deactivate or block a riboswitch. The trigger molecule for a riboswitch (as well as other activating compounds) can be used to activate a riboswitch. Compounds other than the trigger molecule generally can be used to deactivate or block a riboswitch. Riboswitches can also be deactivated by, for example, removing trigger molecules from the presence of the riboswitch. Thus, the disclosed method of deactivating a riboswitch can involve, for example, removing a trigger molecule (or other activating compound) from the presence of the riboswitch. A riboswitch can be blocked by, for example, binding of an analog of the trigger molecule that does not activate the riboswitch.

[0216] Also disclosed are methods for altering expression of an RNA molecule, or of a gene encoding an RNA molecule, where the RNA molecule includes a riboswitch, by bringing a compound into contact with the RNA molecule. Riboswitches function to control gene expression through the binding or removal of a trigger molecule. Thus, subjecting an RNA molecule of interest that includes a riboswitch to conditions that activate, deactivate or block the riboswitch can be used to alter expression of the RNA. Expression can be altered as a result of, for example, termination of transcription or blocking of ribosome binding to the RNA. Binding of a trigger molecule can, depending on the nature of the riboswitch, reduce or prevent expression of the RNA molecule or promote or increase expression of the RNA molecule.

[0217] Also disclosed are methods for regulating expression of an RNA molecule, or of a gene encoding an RNA molecule, by operably linking a riboswitch to the RNA molecule. A riboswitch can be operably linked to an RNA molecule in any suitable manner, including, for example, by physically joining the riboswitch to the RNA molecule or by engineering nucleic acid encoding the RNA molecule to include and encode the riboswitch such that the RNA produced from the engineered nucleic acid has the riboswitch operably linked to the RNA molecule. Subjecting a riboswitch operably linked to an RNA molecule of interest to conditions that activate, deactivate or block the riboswitch can be used to alter expression of the RNA.

[0218] Also disclosed are methods for regulating expression of a naturally occurring gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. If the gene is essential for the survival of a cell or organism that harbors it, activating, deactivating or blocking the riboswitch can in death, stasis or debilitation of the cell or organism. For example, activating a naturally occurring riboswitch in a naturally occurring gene that is essential to survival of a microorganism can result in death of the microorganism (if activation of the riboswitch turns off or represses expression). This is one basis for the use of the disclosed compounds and methods for antimicrobial and antibiotic effects.

[0219] Also disclosed are methods for regulating expression of an isolated, engineered or recombinant gene or RNA that contains a riboswitch by activating, deactivating or blocking the riboswitch. The gene or RNA can be engineered or can be recombinant in any manner. For example, the riboswitch and coding region of the RNA can be heterologous, the riboswitch can be recombinant or chimeric, or both. If the gene encodes a desired expression product, activating or deactivating the riboswitch can be used to induce expression of the gene and thus result in production of the expression product. If the gene encodes an inducer or repressor of gene expression or of another cellular process, activation, deactivation or blocking of the riboswitch can result in induction, repression, or de-repression of other, regulated genes or cellular processes. Many such secondary regulatory effects are known and can be adapted for use with riboswitches. An advantage of riboswitches as the primary control for such regulation is that riboswitch trigger molecules can be small, non-antigenic molecules.

[0220] Also disclosed are methods for altering the regulation of a riboswitch by operably linking an aptamer domain to the expression platform domain of the riboswitch (which is a chimeric riboswitch). The aptamer domain can then mediate regulation of the riboswitch through the action of, for example, a trigger molecule for the aptamer domain. Aptamer domains can be operably linked to expression platform domains of riboswitches in any suitable manner, including, for example, by replacing the normal or natural aptamer domain of the riboswitch with the new aptamer domain. Generally, any compound or condition that can activate, deactivate or block the riboswitch from which the aptamer domain is derived can be used to activate, deactivate or block the chimeric riboswitch.

[0221] Also disclosed are methods for inactivating a riboswitch by covalently altering the riboswitch (by, for example, crosslinking parts of the riboswitch or coupling a compound to the riboswitch). Inactivation of a riboswitch in this manner can result from, for example, that prevents the trigger molecule for the riboswitch from binding, that prevents the change in state of the riboswitch upon binding of the trigger molecule, or that prevents the expression platform domain of the riboswitch from affecting expression upon binding of the trigger molecule.

[0222] Also disclosed are methods for selecting, designing or deriving new riboswitches and/or new aptamers that recognize new trigger molecules. Such methods can involve production of a set of aptamer variants in a riboswitch, assessing the activation of the variant riboswitches in the presence of a compound of interest, selecting variant riboswitches that were activated (or, for example, the riboswitches that were the most highly or the most selectively activated), and repeating these steps until a variant riboswitch of a desired activity, specificity, combination of activity and specificity, or other combination of properties results. Also disclosed are riboswitches and aptamer domains produced by these methods.

[0223] Techniques for in vitro selection and in vitro evolution of functional nucleic acid molecules are known and can

[0224] Also disclosed are methods for selecting and identifying compounds that can activate, deactivate or block a riboswitch. Activation of a riboswitch refers to the change in state of the riboswitch upon binding of a trigger molecule. A riboswitch can be activated by compounds other than the trigger molecule and in ways other than binding of a trigger molecule. The term trigger molecule is used herein to refer to molecules and compounds that can activate a riboswitch. This includes the natural or normal trigger molecule for the riboswitch and other compounds that can activate the riboswitch. Natural or normal trigger molecules are the trigger molecule for a given riboswitch in nature or, in the case of some non-natural riboswitches, the trigger molecule for which the riboswitch was designed or with which the riboswitch was selected (as in, for example, in vitro selection or in vitro evolution techniques). Non-natural trigger molecules can be referred to as unnatural trigger molecules.

[0225] Deactivation of a riboswitch refers to the change in state of the riboswitch when the trigger molecule is not bound. A riboswitch can be deactivated by binding of compounds other than the trigger molecule and in ways other than removal of the trigger molecule. Blocking of a riboswitch refers to a condition or state of the riboswitch where the presence of the trigger molecule does not activate the riboswitch.

[0226] Also disclosed are methods of identifying compounds that activate, deactivate or block a riboswitch. For examples, compounds that activate a riboswitch can be identified by bringing into contact a test compound and a riboswitch and assessing activation of the riboswitch. If the riboswitch is activated, the test compound is identified as a compound that activates the riboswitch. Activation of a riboswitch can be assessed in any suitable manner. For example, the riboswitch can be linked to a reporter RNA and expression, expression level, or change in expression level of the reporter RNA can be measured in the presence and absence of the test compound. As another example, the riboswitch can include a conformation dependent label, the signal from which changes depending on the activation state of the riboswitch. Such a riboswitch preferably uses an aptamer domain from or derived from a naturally occurring riboswitch. As can be seen, assessment of activation of a riboswitch can be performed with the use of a control assay or measurement or without the use of a control assay or measurement. Methods for identifying compounds that deactivate a riboswitch can be performed in analogous ways.

[0227] Identification of compounds that block a riboswitch can be accomplished in any suitable manner. For example, an assay can be performed for assessing activation or deactivation of a riboswitch in the presence of a compound known to activate or deactivate the riboswitch and in the presence of a test compound. If activation or deactivation is not observed as would be observed in the absence of the test compound, then the test compound is identified as a compound that blocks activation or deactivation of the riboswitch.

[0228] Also disclosed are methods of detecting compounds using biosensor riboswitches. The method can include bringing into contact a test sample and a biosensor riboswitch and assessing the activation of the biosensor riboswitch. Activation of the biosensor riboswitch indicates the presence of the trigger molecule for the biosensor riboswitch in the test sample. Biosensor riboswitches are engineered riboswitches that produce a detectable signal in the presence of their cognate trigger molecule. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Biosensor riboswitches can be designed for use in vivo or in vitro. For example, biosensor riboswitches can be designed for use in vivo or in vitro. For example, biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules. Useful biosensor riboswitches can be triggered at or above threshold levels of the trigger molecules.

[0229] Biosensor riboswitches can be used to monitor changing conditions because riboswitch activation is reversible when the concentration of the trigger molecule falls and so the signal can vary as concentration of the trigger molecule varies. The range of concentrations of trigger molecules that can be detected can be varied by engineering riboswitches having different dissociation constants for the trigger molecule. This can easily be accomplished by, for example, “degrading” the sensitivity of a riboswitch having high affinity for the trigger molecule. A range of concentrations can be monitored by using multiple biosensor riboswitches of different sensitivities in the same sensor or assay.

[0230] Also disclosed are compounds made by identifying a compound that activates, deactivates or blocks a riboswitch and manufacturing the identified compound. This can be accomplished by, for example, combining compound identification methods as disclosed elsewhere herein with methods for manufacturing the identified compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound.

[0231] Also disclosed are compounds made by checking activation, deactivation or blocking of a riboswitch by a com-
compound and manufacturing the checked compound. This can be accomplished by, for example, combining compound activation, deactivation or blocking assessment methods as disclosed elsewhere herein with methods for manufacturing the checked compounds. For example, compounds can be made by bringing into contact a test compound and a riboswitch, assessing activation of the riboswitch, and, if the riboswitch is activated by the test compound, manufacturing the test compound that activates the riboswitch as the compound. Checking compounds for their ability to activate, deactivate or block a riboswitch refers to both identification of compounds previously unknown to activate, deactivate or block a riboswitch and to assessing the ability of a compound to activate, deactivate or block a riboswitch where the compound was already known to activate, deactivate or block the riboswitch.

[0232] Disclosed is a method of detecting a compound of interest, the method comprising bringing into contact a sample and a riboswitch, wherein the riboswitch is activated by the compound of interest, wherein the riboswitch produces a signal when activated by the compound of interest, wherein the riboswitch produces a signal when the sample contains the compound of interest. The riboswitch can change conformation when activated by the compound of interest, wherein the change in conformation produces a signal via a conformation dependent label. The riboswitch can change conformation when activated by the compound of interest, wherein the change in conformation causes a change in expression of an RNA linked to the riboswitch, wherein the change in expression produces a signal. The signal can be produced by a reporter protein expressed from the RNA linked to the riboswitch.

[0233] Disclosed is a method comprising (a) testing a compound for inhibition of gene expression of a gene encoding an RNA comprising a riboswitch, wherein the inhibition is via the riboswitch, and (b) inhibiting gene expression by bringing into contact a cell and a compound that inhibited gene expression in step (a), wherein the cell comprises a gene encoding an RNA comprising a riboswitch, wherein the compound inhibits expression of the gene by binding to the riboswitch.

[0234] Also disclosed is a method of identifying riboswitches, the method comprising assessing in-line spontaneous cleavage of an RNA molecule in the presence and absence of a compound, wherein the RNA molecule is encoded by a gene regulated by the compound, wherein a change in the pattern of in-line spontaneous cleavage of the RNA molecule indicates a riboswitch.

A. Identification of Antimicrobial Compounds

[0235] Riboswitches are a new class of structured RNAs that have evolved for the purpose of binding small organic molecules. The natural binding pocket of riboswitches can be targeted with metabolite analogs or by compounds that mimic the shape-space of the natural metabolite. Riboswitches are: (1) found in numerous Gram-positive and Gram-negative bacteria including Bacillus anthracis, (2) fundamental regulators of gene expression in these bacteria, (3) present in multiple copies that would be unlikely to evolve simultaneous resistance, and (4) not yet proven to exist in humans. This combination of features make riboswitches attractive targets for new antimicrobial compounds. Further, the small molecule ligands of riboswitches provide useful sites for derivatization to produce drug candidates.

[0236] Once a class of riboswitch has been identified and its potential as a drug target assessed (by, for example, determining how many genes in a target organism are regulated by that class of riboswitch), candidate molecules can be identified. The following provides an illustration of this using the SAM riboswitch (see Example 7).

[0237] SAM analogs that substitute the reactive methyl and sulfonium ion center with stable sulfur-based linkages (YBD2 and YBD3) are recognized with adequate affinity (low to mid-nanomolar range) by the riboswitch to serve as a platform for synthesis of additional SAM analogs. In addition, a wider range of linkage analogs (N- and C-based linkages) can be synthesized and tested to provide the optimal platform upon which to make amino acid and nucleoside derivations.

[0238] Sulfoxide and sulfone derivatives of SAM can be used to generate analogs. Established synthetic protocols described in Ronald T. Borchardt and Yih Shiong Wu, Potential inhibitor of S-adenosylmethionine-dependent methyltransferase. 1. Modification of the amino acid portion of S-adenosylhomocysteine. J. Med. Chem. 17, 862-868, 1974, can be used, for example. These and other analogs can be synthesized and assayed for binding sequentially or in small groups. Additional SAM analogs can be designed during the progression of compound identification based on the recognition determinants that are established in each round. Simple binding assays can be conducted on B. subtilis and B. anthracis riboswitch RNAs as described elsewhere herein. More advanced assays can also be used.

[0239] The most promising SAM analog lead compounds must enter bacterial cells and bind riboswitches while remaining metabolically inert. In addition, useful SAM analogs must be bound tightly by the riboswitch, but must also fail to compete for SAM in the active sites of protein enzymes, or there is a risk of generating an undesirable toxic effect in the patient’s cells. As a preliminary assessment of these issues, compounds can be tested for their ability to disrupt B. subtilis growth, but fail to affect E. coli cultures (which use SAM but lack SAM riboswitches). To screen for lead compound candidates, parallel bacterial cultures can be grown as follows:

[0240] 1. B. subtilis can be cultured in glucose minimal media in the absence of exogenously supplied SAM analogs.

[0241] 2. B. subtilis can be cultured in glucose minimal media in the presence of exogenously supplied SAM analogs (high doses can be selected, to be followed by repeated experiments designed to test a concentration range of the putative drug compound).

[0242] 3. E. coli can be cultured in glucose minimal media in the presence of exogenously supplied SAM analogs (high doses will be selected, to be followed by repeated experiments designed to test a concentration range of the putative drug compound).

[0243] Fitness of the various cultures can be compared by measurement of cellular doubling times. A range of concentrations for the drug compounds can be tested using cultures grown in microtiter plates and analyzed using a microplate reader from another laboratory. Culture 1 is expected to grow well. Drugs that inhibit culture 2 may or may not inhibit growth of culture 3. Drugs that similarly inhibit both culture 2 and culture 3 upon exposure to a wide range of drug concentrations can reflect general toxicity induced by the exogenous compound (i.e., inhibition of many different cellular processes, in addition to or in place of riboswitch inhibition). Successful drug candidates identified in this screen will
inhibit *E. coli* only at very high doses, if at all, and will inhibit *B. subtilis* at much (>10-fold) lower concentrations.

[0244] As derivatization points on SAM are identified, efficient identification of lead drug compounds will require larger-scale screening of appropriate SAM analogs or generic chemical libraries. A high-throughput screen can be created by one or two different methods using nucleic acid engineering principles. Adaptation of both fluorescent sensor designs outlined below to formats that are compatible with high-throughput screening assays can be accommodated by using immobilization methods or solution-based methods.

[0245] One way to create a reporter is to add a third function to the riboswitch by adding a domain that catalyzes the release of a fluorescent tag upon SAM binding to the riboswitch domain. In the final reporter construct, this catalytic domain can be linked to the yij SAM riboswitch through a communication module that relays the ligand binding event by allowing the correct folding of the catalytic domain for generating the fluorescent signal. This can be accomplished as outlined below.

[0246] SAM Riboswitch Pool Design: A DNA template for in vitro transcription to RNA (FIG. 10) has been constructed by PCR amplification using the appropriate DNA template and primer sequences. In this construct, stem II of the hammerhead (stem PI of the SAM aptamer) has been randomized to present more than 250 million possible sequence combinations, wherein some inevitably will permit function of the ribosyme only when the aptamer is occupied by SAM or a related high-affinity analog. Each molecule in the population of constructs is identical in sequence except at the random domain where multiple copies of every possible combination of sequence will be represented in the population.

[0247] SAM Riboswitch Selection: The in vitro selection protocol can be a repetitive iteration of the following steps:

[0248] 1. Transcribe RNA in vitro by standard methods. Include [α-32P] UTP to incorporate radioactivity throughout the RNA.

[0249] 2. Purify full length RNA on denaturing PAGE by standard methods.

[0250] 3. Incubate full length RNA (~100 pmol) in negative selection buffer containing sufficient magnesium for catalytic activity (20 mM) but no SAM. Incubate 4 h at room temperature (~23°C), with thermocycling or alkaline denaturation as needed to preclude the emergence of selfish molecules.

[0251] 4. Purify full length RNA on denaturing PAGE and discard RNAs that react in the absence of SAM.

[0252] 5. Incubate in positive selection buffer containing 20 mM Mg2+ and SAM (pH 7.5 at 23°C). Incubate 20 min at room temperature.

[0253] 6. Purify cleaved RNA on denaturing PAGE to recover switches that bound SAM and allowed self-cleavage of the RNA.

[0254] 7. Reverse transcribe RNA to DNA.

[0255] 8. PCR amplify DNA with primers that reintroduced cleaved portion of RNA.

[0256] The concentration of SAM in step 4 can be 100 μM initially and can be reduced as the selection proceeds. The progress of recovering successful communication modules can be assessed by the amount of cleavage observed on the purification gel in step 6. The selection endpoint can be either when the population approaches 100% cleavage in 10 mM SAM (conditions for maximal activity of the parental ribosyme and riboswitch) or when the population approaches a plateau in activity that does not improve over multiple rounds. The end population can then be sequenced. Individual communication module clones can be assayed for generation of a fluorescent signal in the screening construct in the presence of SAM.

[0257] A fluorescent signal can also be generated by riboswitch-mediated triggering of a molecular beacon. In this design, riboswitch conformational changes cause a folded molecular beacon tagged with both a fluor and a quencher to unfold and force the fluor away from the quencher by forming a helix with the riboswitch. This mechanism is easy to adapt to existing riboswitches, as this method can take advantage of the ligand-mediated formation of terminator and anti-terminator stems that are involved in transcription control.

[0258] To use riboswitches to report ligand binding by binding a molecular beacon, the appropriate construct must be determined empirically. The optimum length and nucleotide composition of the molecular beacon and its binding site on the riboswitch can be tested systematically to result in the highest signal-to-noise ratio. The validity of the assay can be determined by comparing apparent relative binding affinities of different SAM analogs to a molecular beacon-coupled riboswitch (determined by rate of fluorescent signal generation) to the binding constants determined by standard in-line probing.

**EXAMPLES**

A. Example 1

Coenzyme B1₂ (AdoCbl) Riboswitches

[0259] The example described testing and analysis of a riboswitch that controls gene expression by binding coenzyme B1₂.

[0260] 1. Methods

[0261] i. Chemicals and Oligonucleotides

Coenzyme B1₂ (5'-deoxy-5'-adenosylcobalamin or "AdoCbl") and its analogs methylcobalamin, cobinamide dicyanide, and cyanocobalamin were purchased from Sigma. Trinitated AdoCbl was prepared as described previously (Brown and Zou, Thermolysis of coenzymes B1₂ at physiological temperatures: activation parameters for cobalt-carbon bond homolysis and a quantitative analysis of the perturbation of the homolysis equilibrium by the ribonucleoside triphosphate reductase from *Lactobacillus leichmannii*. J. Inorg. Biochem. 77, 185-195 (1999)). For information regarding the AdoCbl analogs B5,N5'-dimethyl-AdoCbl, N5'-methyl-AdoCbl, N5'-methyl-AdoCbl, N5'-methyl-AdoCbl, N5'-methyl-AdoCbl, PurCbl, 2'-deoxy-AdoCbl and 13-epi-AdoCbl, see Toraya, In: Chemistry and Biochemistry of B1₂, Banerjee, R. Ed. (Wiley, New York) pp. 783-809 (1999).

[0262] DNA oligonucleotides were synthesized by the Keck Foundation Biotechnology Resource Center at Yale University. DNAs were purified by denaturing (8 M urea) PAGE and isolated from the gel by crush soaking in 10 mM Tris-HCl (pH 7.5 at 23°C.), 200 mM NaCl and 1 mM EDTA. The DNA was recovered from the solution by precipitation with ethanol, resuspended in water and stored at -20°C. until use.

[0263] DNA oligonucleotides were synthesized by the Keck Foundation Biotechnology Resource Center at Yale University. DNAs were purified by denaturing (8 M urea) PAGE and isolated from the gel by crush soaking in 10 mM Tris-HCl (pH 7.5 at 23°C.), 200 mM NaCl and 1 mM EDTA. The DNA was recovered from the solution by precipitation with ethanol, resuspended in water and stored at -20°C. until use.

ii. RNA Structure Analysis by In-Line Probing

[0264] Precursor mRNA leader molecules were prepared by in vitro transcription from templates generated by PCR (see In Vivo Expression Constructs and Assays section below) and 5′ 32P-labeled using methods described previously
(Soukop and Breaker, Allostery nucleic acid catalysts. Curr. Opin. Struct. Biol. 10, 318-325 (2000)). Approximately 20 nM of labeled RNA precursor was incubated as described in the brief description of FIG. 1. Accompanying digestions were carried out using reaction conditions similar to those described previously (Soukop and Breaker, Relationship between internucleotide linkage geometry and the stability of P'-NA. RNA 5, 1308-1325 (1999)). To prevent light-induced degradation of ligands, incubations were protected from exposure to light by wrapping each tube with aluminum foil.

[0266] iii. Equilibrium Dialysis Assays

[0267] Each equilibrium dialysis experiment was conducted using a Dispo-Equilibrium Dialyzer (ED-1, Harvard Bioscience) apparatus, wherein two chambers (a and b) each contained 25 μL of equilibration buffer (50 mM Tris-Cl [pH 8.3 at 25°C], 20 mM MgCl₂). The chambers were separated by a dialysis membrane with a 5,000 Dalton molecular weight cut-off. Each experiment (1-IV, boxed), 100 pmol of [³H]-AdoCbl were included in chamber a, and other additives were included as designated (+) for each chamber. In each step, equilibrations were allowed to proceed for 10 hours at 25°C. Before samples were quantitated or before subsequent manipulations were carried out. Quantitation achieved by liquid scintillation counting using 5 or 10 μL of solution from each chamber.

[0268] Dialysis samples were protected from exposure to light by wrapping each apparatus with aluminum foil.

[0269] iv. In Vivo Expression Constructs and Assays

[0270] E. coli K-12 strain was used for all butB-lacZ expression assays and Top10 cells (Invitrogen) were used for plasmid preparation. A DNA (nucleotides ~70 to 450) encompassing the butB leader sequence was amplified as an EcolRI-BamH1 fragment by colony PCR from E. coli strain MC4100 (a gift from S. Gottesman, NIH). The wild-type construct and mutant constructs were inserted into plasmid pRS414 (a gift from R. Simons, UCLA; Simons et al., Improved single and multicopy lac-based cloning vectors for protein and operon fusions, Gene 53, 85-96 (1987)), in-frame with the 9th codon of lacZ (β-galactosidase). Mutant constructs were generated by a three-step PCR strategy wherein regions upstream and downstream of the mutation site were amplified separately with the appropriate DNA primers that introduced the desired sequence changes. The resulting fragments were purified by agarose gel electrophoresis, and then combined and amplified by PCR using primers that correspond to the ends of the full-length construct. The resulting constructs were cloned and sequenced. Constructs whose sequence was confirmed were used for expression analysis and were used as templates for subsequent preparation of PCR-derived DNAs for in vitro transcription.

[0271] The in-frame fusions between various butB leader sequences and lacZ generated as described above were used to determine the levels of expression by employing 3'-ga-lactosidase assay adapted from that described by Miller, In: A Short Course in Bacterial Genetics (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) p. 72 (1992).

[0272] 2. Results

[0273] Metabolite-dependent conformational changes in the 202-nucleotide leader sequence of the butB mRNA. FIG. 1A: Separation of spontaneous RNA-cleavage products of the butB leader using denaturing 10% polyacrylamide gel electrophoresis (PAGE). S-32P-labeled mRNA leader molecules (arrow) were incubated for 41 hr at 25°C in 20 mM MgCl₂, 50 mM Tris-Cl [pH 8.3 at 25°C] in the presence (+) or absence (−) of 20μM of AdoCbl. Lanes containing RNAs that have undergone no reaction, partial digest with alkali, and partial digest with RNase T1 (G-specific cleavage) are identified by NR, −OH, and T1, respectively. The location of product bands corresponding to cleavage after selected guanosine residues are identified by filled arrowheads. Arrowheads labeled 1 through 8 identify eight of the nine locations that exhibit effector-induced structure modulation, which experience an increase or decrease in the rate of spontaneous RNA cleavage. The image was generated using a phosphorimager (Molecular Dynamics), and cleavage yields were quantitated by using ImageQuant software. FIG. 1B: Sequence and secondary-structure model for the 202-nucleotide leader sequence of butB mRNA in the presence of AdoCbl. Putative base-paired elements are designated P1 through P9. Complementary nucleotides in the loops of P4 and P9 that have the potential to form a pseudoknot are juxtaposed. Nine specific sites of structure modulation are identified by light blue arrowheads. The asterisks demark the boundaries of the B1 box (nucleotides 141-162). The coding region and the 38 nucleotides that reside immediately 5' of the start codon (nucleotides 241-243) were not included in the 202-nucleotide fragment. The 315-nucleotide fragment includes the 202-nucleotide fragment, the remaining 38 nucleotides of the leader sequence, and the first 75 nucleotides of the coding region.

[0274] The butB mRNA leader forms a saturable binding site for AdoCbl. FIG. 2A: The dependence of spontaneous cleavage of butB mRNA leader on the concentration of AdoCbl effector as represented by site 1 (G23) and site 2 (U68). S-32P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the in the legend to FIG. 1A, and include identical control and marker lanes as indicated. Incubations contained concentrations of AdoCbl ranging from 10 nM to 100 μM (lunes 1 through 8) or did not include AdoCbl (−). FIG. 2B: Composite plot of the fraction of RNA cleaved at six locations along the mRNA leader versus the logarithm of the concentration (c) of AdoCbl. Fraction cleaved values were normalized relative to the highest and lowest cleavage values measured for each location, including the values obtained upon incubation in the absence of AdoCbl. The inset defines the symbols used for each of six sites, while the remaining three sites were excluded from the analysis due to weak or obscured cleavage bands. Filled and open symbols represent increasing and decreasing cleavage yields, respectively, upon increasing the concentration of AdoCbl. The dashed line reflects a K_c of ~300 nM, as predicted by the concentration needed to generate half-maximal structural modulation. Data plotted were derived from a single PAGE analysis, of which two representative sections are depicted in FIG. 2A.

[0275] The 202-nucleotide mRNA leader causes an unequal distribution of AdoCbl in an equilibrium dialysis apparatus. FIG. 3(I): Equilibration of tritiated effector was conducted in the absence of RNA. FIG. 3(II): (step 1) Equilibration was conducted as in I, but with 200 pmol of mRNA leader added to chamber b; (step 2) 5,000 pmol of unlabeled AdoCbl was added to chamber b. FIG. 3(III): Equilibrations were conducted as described in II, but wherein 5,000 pmol of cyanocobalamin was added to chamber b. IV: (step 1) Equilibration was initiated as described in step 1 of II; (steps 2 and 3) the solution in chamber a was replaced with 25 μL of fresh equilibration buffer; (step 4) 5,000 pmol of unlabeled AdoCbl was added to chamber b. The cpm ratio is the ratio of
counts detected in chamber b relative to that of a. The dashed line represents a cpn ratio of 1, which is expected if equal distribution of tritium is established.

[0276] Selective molecular recognition of effectors by the butB mRNA leader. FIG. 4A shows a chemically structured model of AdoCbl (1) and various effector analogs (2 through 11). FIG. 4B: Determination of analog binding by monitoring modulation of spontaneous cleavage of the 202-nucleotide butB RNA leader. 5'-23P-labeled mRNA leader molecules were incubated, separated, and analyzed as described in the legend to FIG. 1A, and include identical control and marker lanes as indicated. The sections of three PAGE analyses encompassing site 2 (U68) are depicted. Below each image is plotted the amount of RNA cleaved (normalized with relation to the lowest and highest levels of cleavage at U68 in each gel) for each effector as indicated, or for no effector (−). The compound 11 (13-epi-AdoCbl) is an inhibitor of AdoCbl wherein the configuration at C13 is inverted, so that the e propionate side chain is above the plane of the corrin ring; see Brown et al., Conformational studies of 5'-deoxyadenosyl-13-epiocobalamin, a coenzymatically active structural analog of coenzyme B12, Polshyderon 17, 2213 (1998).

[0277] Mutations in the mRNA leader and their effects on AdoCbl binding and genetic control. FIG. 5A: Sequence of the putative P5 element of the wild-type 202-nucleotide butB leader exhibit AdoCbl-dependent modulation of structure as indicated by the observed increase in spontaneous RNA cleavage at position U68 (10% denaturing PAGE gel). Assays were conducted in the absence (−) or presence (+) of 5 nM AdoCbl. The remaining lanes are as described in the legend to FIG. 1A. The composite bar graph reflects the ability of the RNA to shift the equilibrium of AdoCbl in an equilibrium dialysis apparatus and the ability of a reporter gene (see Experimental Procedures) to be regulated by AdoCbl addition to a bacterial culture. (Left) Plotted is the cpn ratio derived by equilibrium dialysis, wherein chamber b contains the RNA. Details of the equilibrium dialysis experiments are described in the brief description of FIG. 3. (Right) Plotted are the expression levels of β-galactosidase as determined from cells grown in the absence (−) or presence (+) of 5 nM AdoCbl. Boxed numbers on the left and right, respectively, reflect the approximate Kd and the fold repression of β-galactosidase activity in the presence of AdoCbl. N.D. designates not determined. FIGS. 5B-5E: Sequences and performance characteristics of various mutant leader sequences as indicated. Constructs were created as described in the Experimental Procedures section.

[0278] i. Metabolite-Induced Structure Modulation of a Messenger RNA.

[0279] To assess whether the butB leader sequence alone is sufficient for sensing and responding to a metabolite, a molecular probing strategy was employed that relies on the structure-dependent spontaneous cleavage of RNA (Soukup and Breaker, Relationship between internucleotide linkage geometry and the stability of P-NA. RNA 5, 1308-1325 (1999); Soukup et al., Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. RNA 7, 524-536 (2001)). The principal mechanism by which an RNA phosphodiester linkage is spontaneously cleaved involves an internal nucleophilic attack by the 2'-oxygen of the adjacent phosphorus center. Since the precise "in-line" positioning of the U-oxygen, phosphorus, and S'-oxygen atoms of a given RNA linkage is essential for a productive nucleophilic attack to occur (Soukup and Breaker, Relationship between internucleotide linkage geometry and the stability of P-NA. RNA 5, 1308-1325 (1999); Soukup et al., Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. RNA 7, 524-536 (2001); Westheimer, Pseudo-rotation in the hydrolysis of phosphate esters. Acc. Chem. Res. 1, 70-78 (1968); Usher, On the mechanism of ribonuclease action. Proc. Natl. Acad. USA 62, 661-667 (1969); Usher and McHale, Hydrolytic stability of helical RNA: a selective advantage for the natural 3',5'-bond. Proc. Natl. Acad. USA 73, 1149-1155 (1976); Descr-Bregeon and Morin, Conformational changes and dynamics of tRNAs: evidence from hydrolysis patterns. Cold Spring Harbor Symp. Quant. Biol. 52, 113-121 (1987)). The rate at which spontaneous cleavage occurs at a given linkage is highly dependent upon the secondary and tertiary structure of the RNA. Specifically, RNA linkages that are formed by nucleotides involved in stable base-paired structures rarely undergo spontaneous cleavage because they rarely adopt an in-line conformation, while nucleotides located in relatively unstructured regions or in tertiary-structured regions experience far greater levels of spontaneous cleavage. Thus, probing of an RNA receptor in the absence and presence of its ligand can be used to provide evidence for RNA structural models and even to determine the dissociation constant for a given RNA-ligand interaction (Soukup and Breaker, Relationship between internucleotide linkage geometry and the stability of P-NA. RNA 5, 1308-1325 (1999); Soukup et al., Generating new ligand-binding RNAs by affinity maturation and disintegration of allosteric ribozymes. RNA 7, 524-536 (2001)).

[0280] A preparation of RNAs that encompass nucleotides 1 through 202 of the 5'-untranslated region of the butB mRNA (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to butB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000); Lundgren et al., Transcribed sequences of the Escherichia coli butB gene control its expression and regulation by vitamin B12. Proc. Natl. Acad. USA 88, 1479-1483 (1991)) was subjected to in-line probing (FIG. 1). In the absence of the putative AdoCbl effector, the RNA exhibits a distinct pattern of cleavage products that is indicative of a well ordered conformational state, which has a mixture of stable structural elements interspersed with regions that are mostly unstructured (FIG. 1A). In the presence of AdoCbl, the pattern of cleavage changes at eight locations, while a ninth position of structural modulation (FIG. 1B) is observed when a longer portion of the mRNA is used. Specifically, metabolite-induced structural modulation at nucleotide 202 (FIG. 1B, position 9) was observed by using in-line probing of a fragment that encompasses nucleotides 1 through 315 of the butB mRNA (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to butB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)). Positions 1, 3, 4, 8, and 9 undergo an effector-dependent dempthing of spontaneous cleavage while the remaining sites experience the reverse effect. A similar pattern of metabolite-modulated RNA cleavage was observed with the analogous 206-nucleotide butB leader RNA of S. typhimurium (Wei et al., Res. Microbiol. 143, 459 (1992)).

[0281] These effector-modulated sites are mapped on a secondary-structure model that was generated by using a combination of computational and RNA probing data. An RNA secondary-structure prediction algorithm (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Bio-
technology (eds. Burciszewski, J., and Clark, B. F. C.) pp. 11-43 (NATO ASI Series, Kluwer Academic Publishers) (1999)) supports a model wherein nine base-paired elements are formed. The in-line probing data and preliminary mutational analyses are consistent with eight of these pairing interactions (P1-P4 and P6-P9), while an alternative pairing interaction (P5) is supported (see below). The majority of these putative base-paired elements appear to remain intact upon effector-induced modulation, with the notable exception of P9. The importance of this structural element in the modulation of ribosome binding and translation has been previously established by mutational analysis (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to btaB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)). Metabolite-dependent formation of the P9 stem-loop structure appears to be critical for the down-regulation of mRNA translation. Consistent with this hypothesis is the observed increase in structure formation in this location upon the addition of AdoCbl (Fig. 1B, decreased cleavage at positions 8 and 9).

A Saturable Metabolite-Binding Site is Formed by a Messenger RNA.

If the structural alteration of the mRNA leader is induced selectively by AdoCbl (as opposed to modulation by a non-specific effect) then the RNA should exhibit characteristics of a typical receptor-ligand interaction. Thus, a plot of the relative extents of structural modulation at each site is expected to yield an apparent dissociation constant (apparent KD) for the effector, which reflects the concentration of effector needed to convert half of the RNAs into their altered structural state. Furthermore, if a single binding event brings about the global structural changes that are observed, then the individual Kr values calculated for each modulation site should converge on a single value, while these values are likely to vary if the structural modulation results from non-specific effects.

Indeed, the levels of spontaneous RNA cleavage were found to correlate with the concentrations of AdoCbl added to the in-line probing mixtures (Fig. 2A). Examination of the dependency of the six most prominent sites of modulation on effector concentration reveals similar apparent K_D values of approximately 300 nM at 25°C. (Fig. 2B). This value is comparable to an apparent K_D value derived from a previous assay that examined the AdoCbl-dependent binding of ribosomes to the btaB mRNA (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to btaB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)). Moreover, the fact that structural modulation occurs over a broad range of concentrations of AdoCbl suggests that this RNA is not likely to make use of cooperative binding of multiple effectors, which would result in a more substantial response to small changes in effector concentration. Together, these observations indicate that the mRNA leader undergoes a substantial change in conformation and forms a high-affinity binding pocket for AdoCbl.

To provide further support for this conclusion, equilibrium dialysis was used to determine whether the RNA could selectively generate an unequal distribution of tritiated AdoCbl (3H-AdoCbl) when incubated in a two-chamber dialysis system. As expected, addition of 3H-AdoCbl to chamber a of an equilibrium dialysis assembly results in near equal distribution of tritium (cpm ratio ~1) between chambers a and b upon incubation (Fig. 3, experiment 1). However, the addition of the 202-nucleotide mRNA leader to chamber b causes a shift in the equilibrium of 3H-AdoCbl (cpm ratio ~2) in favor of chamber b (Fig. 3, experiments II and III). Importantly, the subsequent addition of an excess of unlabeled AdoCbl restores equal distribution of tritium between the two chambers, while the addition of an excess of cyanocobalamin (vitamin B12, an analog of AdoCbl) does not restore the ratio of tritium to unity. Excess unlabeled AdoCbl is expected to restore equal distribution by serving to occupy the vast majority of the binding sites formed by the btaB RNA. In contrast, cyanocobalamin is known to be incapable of serving as a regulatory effector for btaB expression in E. coli (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to btaB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000); Lundrigan and Kadner, Altered cobalamin metabolism in Escherichia coli btaR mutants affects btaB gene regulation. J. Bacteriol. 171, 154-161 (1989)), and thus should be ignored as an effector by the RNA. These findings are consistent with the conclusion that the RNA directly binds AdoCbl and indicate that the RNA forms a selective binding pocket that excludes certain analog compounds.

Assuming that a 1:1 complex is formed between effector and RNA, it was expected that equilibrium dialysis would produce a cpm ratio of far greater than 2 under the assay conditions (2-fold excess RNA over 3H-AdoCbl and concentrations of RNA and effector in excess of the apparent KD). Since there should be an excess of binding sites, the majority of the tritium should be shifted to chamber b upon equilibration. However, the data suggest that ~70% of the tritium in the sample used is not in the form of 3H-AdoCbl. For example, successive replacement of the buffer in chamber a (which removes unshifted tritium from the equilibrium dialysis system) results in increasing values for the cpm ratio (Fig. 3: experiment IV). In addition, the tritium that remains in chamber a upon equilibration with RNA in chamber b cannot be induced to yield an unequal distribution of tritium by btaB RNA in a subsequent equilibrium dialysis experiment (data not shown). The source of this unbound tritium is most likely from light-mediated degradation of AdoCbl, which is highly unstable under ambient light conditions. Mass spectrum analysis of 3H-AdoCbl reveals that the sample is almost entirely intact in the absence of light exposure, but yields ~70% degradation upon exposure to light for a time of about 20 sec) that is typically experienced by a sample when establishing an equilibrium dialysis experiment.

The btaB mRNA Leader Selectively Binds AdoCbl.

To provide selectivity for the genetic response, the btaB mRNA leader must form a precise binding pocket for AdoCbl in order to preclude the genetic switch from being triggered by other metabolites. To explore the molecular recognition capabilities of this RNA, the binding affinity of AdoCbl relative to 10 analogs was indirectly determined (Fig. 4A). This was achieved by determining the extent of spontaneous cleavage at site 2 (nucleotide U68) upon incubation in the presence of AdoCbl or of various analogs (Fig. 4B). It was found that the RNA fails to undergo structural modulation when cobalamin compounds lack the 5'-deoxy-5'-adenosyl moiety. The importance of individual functional groups on this moiety is revealed by the function of other analogs. In summary, modifications at the N1, N3, and N6 positions of the adenine ring cause significant disruption of binding, while the 2'-hydroxyl group of the adjoining ribose moiety is not an important molecular recognition element. Interestingly, a change in the stereochemistry of position 13 of
the corrin ring (compound 11) renders the molecule inactive as a regulatory effector in this in vitro assay and also inside cells. These findings indicate that the butB mRNA leader forms a binding pocket for AdoCbl and that the RNA makes numerous contacts with the effector to ensure high molecular specificity.

[0289] Disruption of Metabolite-RNA Binding has Consequences for Genetic Control.

[0290] The presence of AdoCbl causes reductions in ribosome binding and translation efficiency of the butB mRNA (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to butB mRNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)). The results indicate that this genetic control process is mediated by the selective binding of AdoCbl to the butB mRNA. The effector-binding function of mutant RNA leaders in vitro was compared with their ability to support effector-induced genetic control inside cells. As expected, the wild-type mRNA leader exhibits effector-induced structure modulation, induces an unequal distribution of $^3$H-AdoCbl in an equilibrium dialysis system, and permits down regulation of a reporter gene in E. coli cells treated with AdoCbl and harboring the appropriate reporter construct (summarized in Fig. 5A). However, the introduction of a single mutation (A150T) in the evolutionarily conserved “$I_{12}$ box” (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to butB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)) completely eliminates the in vitro effector-binding and in vivo gene-control functions of this construct, termed “null” (Fig. 5B), which is consistent with the necessity of effector binding for genetic control.

[0291] Mutations that disrupt (U73G, G74U) and subsequently restore (U73G, G74U, C114A, A115C) the predicted P5 stem element were examined. The disrupted stem in construct m2 causes a reduction of AdoCbl binding affinity in vitro and a corresponding reduction in genetic control in cell assays (Fig. 5C), while restoration of the P5 stem element (construct m3) results in near wild-type functions for binding and genetic control (Fig. 5D). This indicates that the P5 stem is an important structural element for function of the RNA. Interestingly, potentially disruptive (m4) and restorative (m5) mutations in a possible pseudoknot structure between the P4 and P9 loops (Fig. IB) both result in a reduction in binding affinity ($K_a \sim 5 \mu M$). If a pseudoknot is being formed, this structure might require a specific sequence for proper function. Although these RNAs maintain diminished but detectable levels of effector binding, neither exhibits genetic control upon the addition of AdoCbl to bacterial cultures harboring the corresponding reporter constructs. The loss in binding affinity likely is sufficient to place these mutant RNAs out of the physiological range for effector concentration, as the cells still retain their natural butB gene whose regulatory system continues to control the import of AdoCbl. The findings support the hypothesis that mRNAs have the structural and functional sophistication needed to perform precision genetic control in the absence of protein regulatory elements.

[0292] v. Analysis

[0293] Genetic control by mRNAs that directly sense the concentrations of metabolites is a newly established paradigm for monitoring the status of cellular metabolism. Although sensing of aminoacyl tRNAs in prokaryotes also appears to be achieved by direct binding of tRNAs to the 5′-untranslated region of their corresponding aminoacyl tRNA synthetases (Henkin, RNA-directed transcription antitermination. Mol. Microbiol. 3, 381-387 (1994)), binding appears to be mediated by Watson/Crick base pairing. In the case of butB the mRNA directly binds the Ado-Cbl effector and becomes resistant to translation initiation, presumably by preventing ribosome binding (Nou and Kadner, Adenosylcobalamin inhibits ribosome binding to butB RNA. Proc. Natl. Acad. Sci. USA 97, 7190-7195 (2000)). If no protein receptors are required for molecular recognition or for modulating gene expression, then this simple “riboswitch” mechanism is most economical in its architecture. Given the organizational simplicity of the butB genetic control components compared to analogous systems that involve proteins, it is likely that mRNAs could be more easily engineered to respond directly to natural and non-biological regulatory effectors.

[0294] It is possible that variations of this mechanism involving direct contacts between metabolite and mRNA are far more widespread in genetic circuitry. For example, the S. typhimurium cob operon, which encodes proteins in the biosynthetic pathway for the AdoCbl coenzyme, carries $B_{12}$ box and other regulatory structures in its leader domain (Ravnum and Andersson, An adenosylcobalamin (coenzyme-$B_{12}$)-repressed translational enhancer in the cob mRNA of Salmonella typhimurium. Mol. Microbiol. 39, 1585-1594 (2001)). It has been noted (White III, Coenzymes as fossils of an earlier metabolic state. J. Mol. Evol. 7, 101-104 (1976)) that these two coenzymes and FMN, which is another potential riboswitch effector (Gieland et al., A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. Trends Genetics 15, 439-442 (1999)), possibly are molecular fossils of an ancient metabolic state that was run entirely by RNA. If true, then mechanisms involving metabolite sensing by mRNA might be one of the oldest forms of genetic control in existence.

B. Example 2
Thiamine Pyrophosphate (TPP) Riboswitches

[0295] The example described testing and analysis of a riboswitch that controls gene expression by binding thiamine pyrophosphate.

[0296] 1. Chemicals and Oligonucleotides

[0297] TPP, thiamine monophosphate (TP), thiamine, oxythiamine, amprolium, and benfotiamine were purchased from Sigma. Thiamine disulfide and 4-methyl-5-[3]-hydroxyethylthiazole (THZ) were purchased from TCI America. $^3$H-labeled thiamine was purchased from American Radiolabeled Chemicals, Inc. (10 Ci mmol$^{-1}$). Synthetic DNAs were synthesized by the Keck Foundation Biotechnology Resource Center at Yale University. DNAs were purified by denaturing (8 M urea) polyacrylamide gel electrophoresis (PAGE) and isolated from the gel by crush-soaking in 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl and 1 mM EDTA. The DNA was recovered by precipitation with ethanol.

[0298] 2. Construction of E. coli thiM- and E. coli thiC-lacZ Fusions

[0299] Nucleotides 83 to 238 of the E. coli thiCFeGH operon (Vander Horn et al., Structural genes for thiamine biosynthetic enzymes (thiCFeGH) in Echerichia coli K-12. J. Bacteriology 175, 982-992 (1993), was amplified by PCR from E. coli strain MC4100 (obtained from S. Gottesman, NIH) as a EcoRI-Bgl II fragment. The DNA was ligated into EcoR1- and BamHI-digested pBR341 plasmid DNA, which contains a promoterless copy of lacZ (obtained from R. Simons, UCLA; Simons et al., Improved single and multi-
copy lac-based cloning vectors for protein and operon fusions (Gene 53, 85-96 (1987)), resulting in the in-frame fusion of the 9th codon of lacZ to the 9th codon of thiC. Similarly, the regulatory region of thiM (nucleotides -67 to 163) was amplified by PCR as a EcoR1-BamH1 fragment and inserted into plasmid pRS414, wherein the 6th codon of thiM resides in-frame with the 9th codon of lacZ. The plasmids were transformed into Top10 cells (Invitrogen) for all subsequent manipulations. All site-directed mutations were introduced into the thiC and thiM regulatory regions using the QuickChange site-directed mutagenesis kit (Stratagene) and the appropriate mutagenic DNA primers. All mutations were confirmed by DNA sequencing (USB Thermosequenase).

[0300] 3. Thiamine-Repression β-Galactosidase Assays
[0301] E. coli cells (Top10; Invitrogen) that contained in-frame lacZ fusions to thiC or thiM mRNA leader sequences, were grown in M9 glucose minimal media (plus 50 μg/ml Vitamin assay Casamino acids; Difco) to mid-exponential phase. The cultures were either grown with or without added thiamine (100 μM). Aliquots (1 ml) were removed for β-galactosidase enzyme assays, which were conducted in a manner similar to that described by Miller (Miller, In: A Short Course in Bacterial Genetics Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., p. 72, (1992)). All assays were repeated twice and in duplicate, with Miller unit values reflecting the average of these analyses.

[0302] 4. In Vitro Transcription
[0303] Templates for in vitro transcription of the fragments of thiC and thiM mRNA leaders were generated by PCR using the appropriate DNA primers and plasmids pRS414thiC or pRS414thiM, respectively. The dinucleotide sequence GG was introduced into the DNA constructs (corresponding to the 5’ terminus of each RNA construct) at this step to facilitate transcription by T7 RNA polymerase. RNAs were prepared by in vitro transcription and were 5’ 32P-labeled as described previously (Seetharaman et al., Immobilized riboswitches for the analysis of complex chemical and biological mixtures. Nature Biotechnol. 19, 336-341 (2001)).

[0304] 5. In-Line Probing of RNA
[0305] Determination of apparent Kd values for each construct was achieved by conducting in-line probing of RNA constructs wherein the concentration of the ligand varied between 10 nM and 10 μM, or up to 10 μM for weakly binding ligands. Specifically, TPP-dependent modulation of the spontaneous cleavage of RNA constructs was visualized by polyacrylamide gel electrophoresis (PAGE). 5’ 32P-labeled RNAs (20 nM) were incubated for approximately 40 hr at 25°C in 20 mM MgCl2, 50 mM Tris-HCl (pH 8.3 at 25°C) in the presence (+) or absence (−) of 100 μM TPP. Some RNAs were subjected to no reaction, partial digestion with alkali, or partial digestion with RNase T1 (G-specific cleavage) (see FIG. 6a). Composite plots of the fraction of RNA cleaved at specific sites versus the logarithm of the concentration of ligand (e.g. FIG. 7a) were generated to provide an estimate of the apparent Kd. Fraction cleaved values were normalized relative to the highest and lowest cleavage values measured for each site.

[0306] 6. Equilibrium Dialysis
[0307] Equilibrium dialysis assays were conducted using a DispoEquilibrium Dialyzer (ED-1, Harvard Bioscience), wherein chambers a and b were separated by a 5,000 Dalton molecular weight cut-off membrane. Equilibration was initiated by the addition of 25 μl of equilibration buffer [50 mM Tris-HCl (pH 8.3 at 25°C), 20 mM MgCl2, 100 mM KCl], containing 100 nM 3H-thiamine and by the addition of an equal volume of equilibration buffer either without or with 20 μM RNA as indicated to chamber b. Equilibrations were allowed to proceed for 10 hr at 23°C, and aliquots were removed from each chamber and quantitated by using a liquid scintillation counter.

[0308] 7. Results
[0309] i. Metabolite Binding by mRNAs.

[0310] FIG. 6A shows TPP-dependent modulation of the spontaneous cleavage of 165 thiM RNA was visualized by polyacrylamide gel electrophoresis (PAGE). 5’ 32P-labeled RNAs (arrow, 20 nM) were incubated for approximately 40 hr at 25°C in 20 mM MgCl2, 50 mM Tris-HCl (pH 8.3 at 25°C) in the presence (+) or absence (−) of 100 μM TPP. NR, OH1 and T1 represent RNAs subjected to no reaction, partial digestion with alkali, or partial digestion with RNase T1 (G-specific cleavage), respectively. Product bands representing cleavage after selected G residues are numbered and identified by filled arrowheads. The asterisk identifies modulation of RNA structure involving the Shine-Dalgarno (SD) sequence. Gel separations were analyzed using a phosphorimagery (Molecular Dynamics) and quantitated using ImageQuant software.

[0311] FIG. 6B shows a secondary-structure model of 165 thiM as predicted by computer modeling (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Biotechnology (eds. Barciszewski J. & Clark, D. F. C.) 11-43 (NATO ASI Series, Kluwer Academic Publishers, 1999); Mathews et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911-940 (1999)) and by the structure probing data depicted in FIG. 6A. Spontaneous cleavage characteristics are as noted in the inset. Unmarked nucleotides exhibit a constant but low level of degradation. The truncated 91 thiM RNA is boxed and the thi box element (Miranda-Rios et al., A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc. Natl. Acad. Sci. USA 98, 9736-9741 (2001)) is shaded. Nucleotides enclosed in boxes identify an alternative pairing, designated P8°. The RNA carries two mutations (G156A and U157C) relative to wild type that were introduced in a non-essential portion of the construct to form a restriction site for cloning, while all RNAs carry two 5’ terminal G residues to facilitate in vitro transcription.

[0312] FIG. 6C shows TPP-dependent modulation of the spontaneous cleavage of 240 thiC RNA. Reactions were conducted and analyzed as described in above for FIG. 6A. FIG. 6D shows a secondary-structure model of 240 thiC. Base-paired elements that are similar to those of thiM are labeled P1 through P5. The truncated RNA 111 thiC is boxed. Nucleotides enclosed in boxes identify an alternative pairing.

[0313] ii. The thiM and thiC mRNA Leaders Serve as High-Affinity Metabolite Receptors.

[0314] FIG. 7A shows the extent of spontaneous modulation of RNA cleavage at several sites within 165 thiM (left) and 240 thiC (right) plotted for different concentrations (c) of TPP. Arrows reflect the estimated concentration of TPP needed to attain half maximal modulation of RNA (apparent Kd). FIG. 7B shows the logarithm of the apparent Kd values plotted for both RNAs with TPP, TP and thiamine as indicated. The boxed data was generated using TPP with the truncated RNAs 91 thiM and 111 thiC. FIG. 7C shows that patterns of spontaneous cleavage of 165 thiM differ between
thiamine and TPP ligands as depicted by PAGE analysis (left) and as reflected by graphs (right) representing the relative phosphorimager counts for the three lanes as indicated. Details for the RNA probing analysis are similar to those described above in connection with FIG. 6A. The graphs were generated by ImageQuant software. [0315] iii. High Sensitivity and Selectivity of mRNA Leaders for Metabolite Binding. [0316] FIG. 8A shows chemical structures of several analogues of thiamine. TD is thiamine disulfide and THZ is 4-methyl-5-β-hydroxyethylthiazole. FIG. 8B shows PAGE analysis of 165 thm RNA structure probing using TPP and various chemical analogues (40 µM each) as indicated. Locations of significant structural modulation within the RNA spanning nucleotides ~113 to ~159 are indicated by open arrowheads. The asterisk identifies the site (C144) used to compare the normalized fraction of RNA that is cleaved (bottom) in the presence of specific compounds. Details for the RNA probing analysis are similar to those described above in connection with FIG. 6A. FIG. 8C shows a summary of the features of TPP that are critical for molecular recognition. FIG. 8D shows equilibrium dialysis using 3H-thiamine as a tracer. Plotted are the ratios for tritium distribution in a two-chamber system (a and b) that were established upon equilibration in the presence of the RNA constructs in chamber b as indicated (see below for a description of the non-TPP-binding mutant M3). 100 µM TPP or oxythiamine were added to chamber a, as denoted, upon the start of equilibration. [0317] iv. Mutational Analysis of the Structure and Function of the thm Riboswitch. [0318] FIG. 9A shows mutations present in constructs M1 through M8 relative to the 165 thm RNA. P8* is a putative base-paired element between portions (shaded) of the P1 and P8 stems. FIGS. 9B and 9C show in vitro ligand-binding and genetic control functions of the wild-type (WT), M1 and M2 RNAs as reflected by PAGE analysis of in-line probing experiments (10 µM TPP) and by β-galactosidase expression assays. Labels on PAGE gels are as described above in connection with FIG. 6A. Bars represent the levels of gene expression in the presence (+) and the absence (−) of TPP in the culture medium. FIG. 9D is a summary of similar analyses of WT through M9 is presented in table form. The SD status “n.d.” (not determined) indicates either that the level of spontaneous cleavage detected in the absence and presence of TPP is near the limit of detection (M6, M7 and M8) or that the region adopts an unusual structure (M9) compared to WT. [0319] 8. Discussion [0320] β-galactosidase fusion constructs were prepared that encompass the 5′-untranslated region of thm and thiC mRNAs of E. coli, which includes a previously identified “thi box” domain whose sequence and potential secondary structure are conserved in several species of bacteria and archaea (Miranda-Rios et al., A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc. Natl. Acad. Sci. USA 98, 9736-9741 (2001)). The thm and thiC translational fusion constructs exhibit thiamine-dependent repression of β-galactosidase activity of 18- and 110-fold, respectively, when host cells are grown in a minimal medium that otherwise lacks a source of thiamine. A transcriptional fusion containing the thiM leader is not subject to suppression by thiamine, but a similar fusion with thiC leader yields a 16-fold modulation with thiamine, suggesting that a significant portion of genetic control observed with thiC occurs at the level of transcription. [0321] These constructs were used to prepare DNA templates by PCR for in vitro transcription of RNA fragments. The resulting RNAs were subjected to a structure-probing process (see Example 1) to reveal whether the RNAs undergo structure modulation upon binding of ligands. Interヌucleotide linkages in unstructured regions are more likely to undergo spontaneous cleavage compared to linkages that reside in highly structured regions of an RNA (Scoukup & Breaker, Relationship between internucleotide linkage geometry and the stability of RNA. RNA 5, 1308-1325 (1999)). The 165-nucleotide thim RNA fragment (165 thm) has a distinct pattern of cleavage products that is generated when the RNA is incubated for an extended period in the absence of TPP (FIG. 6A). Upon addition of 100 µM TPP, 165 thim undergoes substantial structural alteration as many internucleotide linkages in the region spanning positions 39 through 80 exhibit a reduction in spontaneous cleavage. This indicates that TPP binds to the RNA and stabilizes a defined structure within this region, resulting in a lower rate of fragmentation. [0322] The fragmentation patterns are largely congruent with potential base-paired and bulge structures that are identified by a secondary-structure prediction algorithm (Zuker et al., Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In RNA Biochemistry and Biotechnology (eds. Barciszewski J. & Clark, B. F. C.) 11-43 (NATO ASI Series, Kluwer Academic Publishers, 1999); Mathews et al., Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911-940 (1999)). Most linkages that experience a ligand-induced reduction of cleavage are encompassed by the thi box and nucleotides that reside immediately 5′ relative to this domain (FIG. 6B). Other linkages that undergo cleavage, but that are not modulated by TPP, are predicted to reside in bulges or in loops of hairpins. Predicted base-paired structures labeled P2 through P7 encompass linkages that exhibit the lowest levels of spontaneous cleavage, implying that they remain structured in both the presence and absence of TPP. Interestingly, nucleotides 126 through 130 encompass the only region apart from those described above that become more structured upon TPP addition. These nucleotides correspond to the Shine-Dalgarno (SD) sequence, which is required for efficient translation of mRNAs in prokaryotes. These findings are consistent with a genetic control mechanism wherein the thm RNA binds to TPP and forms a complex wherein the ribosome cannot gain access to the SD sequence. [0323] Similarly, structure probing was used to examine the mRNA leader for thiC. The 240 thiC RNA also exhibits extensive modulation of its pattern of spontaneous cleavage, and again the majority of the changing pattern is located in the thi box and in the region located immediately upstream of this domain (FIG. 6C). These regions of highest structure modulation in thm and thiC can be folded into similar secondary structures (FIG. 6D), and carry several common sequence elements within and adjacent to the thi box domain. Thus, the structures of thm and thiC spanning stems P1 through P5 comprise TPP-binding motifs that are analogous to aptamers, which are engineered ligand-binding RNAs (Osborne & Ellington, Nucleic acid selection and the challenge of combinatorial chemistry. Chem. Rev. 97, 349-370 (1997); Hermann & Patel, Adaptive recognition by nucleic acid aptamers. Science 287, 820-825 (2000); Gold et al., Diversity of oligonucleotide aptamers. Annu. Rev. Biochem. 64, 763-797)
(1995)). Nucleotides residing 3' relative to this natural TPP aptamer are involved in converting the metabolite binding event into a genetic response.

[0324] The sensitivity of metabolite detection by these mRNAs was assessed by establishing apparent dissociation constant (apparent K_p) values for TPP, thiamine, and thiamine monophosphate (TP). Values were generated by monitoring the extent of spontaneous cleavage at several ligand-sensitive sites within the RNA under a range of ligand concentrations. For example, probing of a trace amount of 165 thiM RNA under TPP concentrations ranging from zero to 100 μM (or up to 10 mM with certain analogues) reveals that half-maximal modulation of RNA structure occurs when approximately 600 nM TPP is present (FIG. 7A), which reflects an apparent K_p of 600 nM. Likewise, probing of 240 thiC reveals an apparent K_p of 100 nM. Both 165 thiM and 240 thiC RNAs appear to bind TPP more avidly than TP or thiamine, with thiC exhibiting more than 1,000-fold discrimination against TP and thiamine (FIG. 7B). The fact that TPP is the strongest modulator of RNA structure is consistent with genetic observations in Salmonella typhimurium that TPP synthesis is required for regulation of expression of thiamine biosynthesis genes (Webb et al., Thiamine pyrophosphate (TPP) negatively regulates transcription of some thi genes of Salmonella typhimurium. J. Bacteriol. 178, 2533-2538 (1996)). The differential specificity achieved by the RNAs, which is a phenomenon that is commonly observed for receptor-ligand systems made of protein, indicates that these ligand-binding RNAs would be receptive to specific changes (through, for example, natural or artificial evolutionary forces).

[0325] The actual K_p values for RNA-ligand interactions might be different inside cells where physiological conditions of Mg^{2+} and other agents that can influence RNA structure will not match those of the in vitro assays. Also, the nature of the RNA construct can be a source of an altered K_p. For example, the minimized 91 thiM construct (FIG. 6A), which largely encompasses only the putative natural aptamer, retains the ability to bind TPP and exhibits an apparent K_p that is improved by approximately 20 fold compared to the longer construct (FIG. 7B). Thus, the affinity for TPP might vary as the nascent RNA transcript emerges from the active site of RNA polymerase or the ribosome. Furthermore, this result demonstrates that the 91 thiM aptamer domain can be separated from RNA components (collectively termed the “expression platform”) that are directly controlling gene expression. This modular construction, involving the physical and functional separation of aptamer and expression platform domains allows the generation of TPP-controlled RNAs by rational RNA engineering strategies (or through evolutionary processes).

[0326] Spontaneous cleavage at several linkages within the thi box domain of 165 thiM specifically correlate with the type of ligand used. Although TPP reduces spontaneous cleavage of 165 thiM at nucleotides A61, U62 and to a smaller extent at U79, these same sites retain an elevated level of cleavage when thiamine is present near its saturating concentration (FIG. 7C). These nucleotides cluster at an internal bulge within the thi box domain, and appear to contribute to the binding site for the phosphate groups of TPP.

[0327] The structural modulation of 165 thiM was further examined in the presence of several analogues that carry certain structural features of thiamine (FIG. 8A). Thiamine and its phosphorylated derivatives TP and TPP induce modulation as expected (FIG. 8B). However, oxthiamine and other thiamine analogues with less similarity to TPP fail to induce structure modulation. The performance of this sampling of analogues indicates that the RNA makes specific contacts to distal parts of its ligand and that both the purine and phosphate groups carry important elements for molecular recognition (FIG. 8C). Similar results are obtained by using equilibrium dialysis assays (FIG. 8D). For example, the addition of 91 thiM RNA to chamber b of an equilibrium dialysis assembly causes a shift in the distribution of 3^H-thiamine in favor of chamber b, unless an excess of unlabeled TPP is also included. However, the presence of oxthiamine does not significantly restore the tritium distribution to unity, which is expected because probing data indicate that it is not able to bind the RNA. These findings indicate that the aptamer domain of the TPP riboswitch is highly selective for its target ligand.

[0328] The secondary structure model for 165 thiM RNA was examined in greater detail by generating and testing a series of variant constructs (FIG. 9A). For example, variant M1 carries a mutation that disrupts the predicted P3 pairing element. This mutation causes a loss of TPP binding (FIG. 9B, e.g., see position C77) and a loss of genetic control of the corresponding β-galactosidase fusion construct (FIG. 9C, graph). Re-establishment of base pairing in the double-mutant construct M2 restores both TPP binding and genetic control. Similarly, disruptive and restorative mutations encompassed by constructs M3 through M5 are consistent with the formation of stems P5 and P8. Upon the addition of TPP, the SD element of both the WT and M2 constructs becomes sequestered in a structure that precludes a high level of spontaneous cleavage. In contrast, the M1 construct does not exhibit SD modulation (FIGS. 9B and 9C, nucleotides 126-130). These results are consistent with the genetic switch being turned off by a mechanism whereby TPP binding ultimately promotes the stable formation of P8, which reduces access to the SD by the ribosome.

[0329] The partner of the SD sequence in P8 (nucleotides 108 to 111) remains resistant to spontaneous cleavage both in the presence and absence of TPP (FIG. 6A). This is consistent with the formation of P8, upon addition of TPP, due to the displacement of an alternative structure that otherwise prevents this anti-SD element from forming P8. Furthermore, nucleotides 83 through 86 are complementary to the anti-SD element and this region also resists spontaneous cleavage in the presence and absence of TPP. A mechanism by which genetic control could result, which is tested as described above, is via the mutually exclusive formation of P8* in the ‘On’ state versus the simultaneous formation of P1 and P8 in the metabolite-bound ‘Off’ state (FIG. 9D).

[0330] Constructs M7 through M9 were tested in an assessment of this mechanism. Construct M7 carries a U109C mutation in the anti-SD sequence that is designed to destabilize the P8 interaction while simultaneously destabilizing the P8* interaction. M7 retains TPP binding function and exhibits a significant level of generic modulation (FIG. 9C, box), which is expected if the mutation does not disrupt the relative distribution of mRNAs between the ‘On’ and ‘Off’ states. In comparison, M8 (U110C) retains TPP binding, exhibits a dramatic reduction in the level of reporter expression, and loses nearly all genetic modulation. In addition, M8 no longer exhibits detectable spontaneous cleavage in the SD sequence, which is consistent with the thermodynamic balance between P8 and P8* formation being shifted decidedly in favor of P8
in this RNA variant. Construct M9, which carries four mutations in the anti-SD element, has a significantly different pattern of spontaneous cleavage in the SD region. M9 fails to reduce gene expression upon thiamine addition to cells, despite the fact that the construct retains TPP binding in vitro. It is evident from these data that TPP binding restricts the structural freedom of the SD element in the appropriate RNA variants, and that this correlates with genetic control.

C. Example 3

Metabolite-Binding Riboswitches

1. Introduction

Modern organisms must coordinate the expression of many hundreds of genes in response to metabolic demands and environmental changes. Each gene product must be regulated temporally, quantitatively, and oftentimes spatially. Additionally, genetic control processes must be dynamic, rapid, and selectively responsive to the specific conditions undergoing change. Therefore, organisms require sentries of genetic regulatory factors that continuously quantify a multitude of environmental signals. Upon measurement of a particular signal, which may be one of many possible biochemical or physical cues, these regulatory factors must modulate expression of a specific subset of the organism’s genes.

It has generally been assumed that proteins are the obligate sensors of these signals because proteins are a proven medium for forming highly responsive sensors. However, it was discovered that mRNAs also are capable of acting as direct sensors of chemical and physical conditions for the purpose of genetic control. Classes of mRNA domains, collectively referred to as ‘riboswitches’, serve as RNA genetic control elements that sense the concentrations of specific metabolites by directly binding the target compound. Riboswitches that have been discovered are responsible for sensing metabolites that are critical for fundamental biochemical processes including adenosylcobalamin (AdoCbl) (see Example 1), thiamine pyrophosphate (TPP) (see Example 2), flavin mononucleotide (FMN), S-adenosylmethionine (SAM) (see Example 7), tRNA (see Example 5), guanine (see Example 6), and adenine (see Example 8). Upon interaction with the appropriate small molecule ligand, riboswitch mRNAs undergo a structural reorganization that results in the modulation of genes that they encode. To date, all riboswitches that have been examined in detail cause genetic repression upon binding their target ligand, although riboswitches that activate gene expression upon ligand binding can be produced (and will likely be found in nature).

Each instance, riboswitch domains have been subjected to a battery of biochemical and genetic analyses in order to convincingly demonstrate that direct interaction of small organic metabolites with mRNA receptors leads to a corresponding alteration in genetic expression. This example provides a brief summary of these efforts and of some of the general characteristics that are exhibited by riboswitches. Using these discoveries and the principles of riboswitch operation described in this example and elsewhere herein, those of skill in the art can use and adapt riboswitches for many purposes including use as genetic tools and as targets for development of antimicrobials.

2. General Organization of Riboswitch RNAs

Bacterial riboswitch RNAs are genetic control elements that are located primarily within the 5' untranslated region (5'-UTR) of the main coding region of a particular mRNA. Structural probing studies (discussed further below) revealed that riboswitch elements are generally composed of two domains: a natural aptamer (T. Hermann, D. J. Patel, Science 2000, 287, 820; L. Gold, et al., Annual Review of Biochemistry 1995, 64, 763) that serves as the ligand-binding domain (referred to herein as the aptamer domain), and an ‘expression platform’ that interfaces with RNA elements that are involved in gene expression (e.g. Shine-Dalgarno (SD) elements; transcription terminator stems). These conclusions are drawn from the observation that aptamer domains synthesized in vitro bind the appropriate ligand in the absence of the expression platform (see Examples 2 and 6). Moreover, structural probing investigations suggest that the aptamer domain of most riboswitches adopts a particular secondary- and tertiary-structure fold when examined independently, that is essentially identical to the aptamer structure when examined in the context of the entire 5' leader RNA. This implies that, in many cases, the aptamer domain is a modular unit that folds independently of the expression platform (see Examples 2 and 6).

Ultimately, the ligand-bound or unbound status of the aptamer domain is interpreted through the expression platform, which is responsible for exerting an influence upon gene expression. The view of a riboswitch as a modular element is further supported by the fact that aptamer domains are highly conserved amongst various organisms (and even between kingdoms as is observed for the TPP riboswitch, whereas the expression platform varies in sequence, structure, and in the mechanism by which expression of the appended open reading frame is controlled. For example, ligand binding to the TPP riboswitch of the thiamin mRNA of B. subtilis causes transcription termination. This expression platform is distinct in sequence and structure compared to the expression platform of the TPP riboswitch in the thiM mRNA from E. coli, wherein TPP binding causes inhibition of translation by a SD blocking mechanism (see Example 2). The TPP aptamer domain is easily recognizable and of near identical functional character between these two transcriptional units, but the genetic control mechanisms and the expression platforms that carry them out are very different.

Aptamer domains for riboswitch RNAs typically range from ~70 to 170 nt in length (FIG. 11). This observation was somewhat unexpected given that in vitro evolution experiments identified a wide variety of small molecule-binding aptamers, which are considerably shorter in length and structural intracacy (T. Hermann, D. J. Patel, Science 2000, 287, 820; L. Gold, et al., Annual Review of Biochemistry 1995, 64, 763; M. Famulok, Current Opinion in Structural Biology 1999, 9, 324). The substantial increase in complexity and information content of the natural aptamer sequences relative to artificial aptamers is most likely required to form RNA receptors that function with high affinity and selectivity. Apparent K_d values for the ligand-riboswitch complexes range from low nanomolar to low micromolar. It is also worth noting that some aptamer domains, when isolated from the appended expression platform, exhibit improved affinity for the target ligand over that of the intact riboswitch (~10 to 100-fold) (see Example 2). This likely represents an energetic cost in sampling the multiple distinct RNA conformations required by a fully intact riboswitch RNA, which is reflected by a loss in ligand affinity. Since the aptamer domain must serve as a molecular switch, this might also add to the functional demands on natural aptamers that might help rationalize their more sophisticated structures.
3. Riboswitch Regulation of Transcription Termination in Bacteria

Bacteria primarily make use of two methods for termination of transcription. Certain genes incorporate a termination signal that is dependent upon the Rho protein (J. P. Richardson, Bioclinica et Biophysica Acta 2002, 1577, 251), while others make use of Rho-independent terminators (intronic terminators) to destabilize the transcription elongation complex (I. Gusarov, E. Nudler, Molecular Cell 1999, 3, 495; E. Nudler, M. E. Gottesman, Genes to Cells 2002, 7, 755). The latter RNA elements are composed of a GC-rich stem-loop followed by a stretch of 6-9 uridyl residues. Intronic terminators are widespread throughout bacterial genomes (F. Lillo, et al., Bioinformatica 2002, 18, 971), and are typically located at the 3’-termini of genes or operons. Interestingly, an increasing number of examples are being observed for intronic terminators located within 5’-UTRs.

Amongst the wide variety of genetic regulatory strategies employed by bacteria there is a growing class of examples wherein RNA polymerase responds to a termination signal within the 5’-UTR in a regulated fashion (T. M. Henkin, Current Opinion in Microbiology 2003, 3, 149). During certain conditions the RNA polymerase complex is directed by external signals either to perceive or to ignore the termination signal. Although transcription initiation might occur without regulation, control over mRNA synthesis (and of gene expression) is ultimately dictated by regulation of the intrinsic terminator. Generally, one of at least two mutually exclusive mRNA conformations results in the formation or disruption of the RNA structure that signals transcription termination. A trans-acting factor, which in some instances is a RNA (F. J. Grundy, et al., Proceedings of the National Academy of Sciences of the United States of America 2002, 99, 11121; T. M. Henkin, C. Yanofsky, Bioessays 2002, 24, 700) and in others is a protein (J. Stulke, Archives of Microbiology 2002, 177, 433), is generally required for receiving a particular intracellular signal and subsequently stabilizing one of the RNA conformations. Riboswitches offer a direct link between RNA structure modulation and the metabolite signals that are interpreted by the genetic control machinery. A brief overview of the FMN riboswitch from a B. subtilis mRNA is provided below to illustrate this mechanism.

i. A Natural Aptamer for FMN

A highly conserved RNA domain, referred to as the RFN element, was identified in bacterial genes involved in the biosynthesis and transport of riboflavin and FMN (M. S. Gelfand, et al., Trends in Genetics 1999, 15, 439; A. G. Vitreschak, et al., Nucleic Acids Research 2002, 30, 3141). This element is required for genetic manipulation of the riboflavin operon (hereafter, “rifD”) of B. subtilis, as mutations resulted in a loss of FMN-mediated regulation (Y. V. Kil, et al., Molecular & General Genetics 1992, 233, 483; V. N. Mironov, et al., Molecular & General Genetics 1994, 242, 201). These data led to the proposal that either a protein-based sensor, or FMN itself (G. D. Stormo, Y. Ji, Proceedings of the National Academy of Sciences of the United States of America 2001, 98, 9465) interacts with the RFN element in order to repress rifD gene expression. However, there was no understanding of how such interactions would take place or the mechanism by which expression would be affected. Although RNA sequences that specifically bind FMN had been identified through directed evolution experimentation (C. T. Lauhon, J. W. Szostak, Journal of the American Chemical Society 1995, 117, 1246, M. Roychowdhury-Saha, et al., Biochemistry 2002, 41, 2492), they exhibit no obvious resemblances to the RFN element.

ii. Structural Probing Reveals FMN-Mediated RNA Structure Modulation

Each internucleotide linkage in a RNA polymer is susceptible to spontaneous hydrolysis by an S2-like mechanism, wherein the 2’ oxygen attacks the adjacent phosphorus center, leading to chain cleavage. This reaction requires a 180° orientation between the attacking nucleophile, the phosphorus center, and the 2’-oxygen leaving group (in-line conformation) (G. A. Soukup, R. R. Breakey, RNA 1993, 3, 1308; V. Tereshko, et al., RNA 2001, 7, 405). Nucleotides that are base-paired, or otherwise structurally constrained, are typically incapable of adopting this configuration and therefore display low rates of spontaneous cleavage. In contrast, nucleotides that are structurally unrestrained exhibit much higher rates of spontaneous cleavage. These observations have been exploited in a structural probing method, referred to as “in-line probing”, which establishes the relative rates of spontaneous cleavage for a given RNA polymer and correlates this with secondary- and tertiary-structure models (V. Tereshko, et al., RNA 2001, 7, 405).

To assess whether the RFN element of rifD was sensitive to FMN, a fragment of the corresponding 5’-UTR was 5’-32P labeled and incubated in the absence and presence of FMN, and the resulting fragments were analyzed by polyacrylamide gel electrophoresis (PAGE). Interestingly, patterns differ between reactions with and without FMN, signifying that there is a structural rearrangement of the RNA upon FMN binding to rifD. The spontaneous cleavages of certain nucleotide positions located within inter-helical regions of the RFN element become significantly reduced in the presence of FMN, suggesting that these nucleotides are involved in forming an FMN-RNA complex, which forces structural constraints upon the RNA (FIG. 12). It is of this type of structural modulation that can be harnessed by the expression platform for allosteric modulation of gene expression.

Additional evidence for direct binding of FMN by the rifD RFN element was generated by enzymatic probing. Oligonucleotides predicted to anneal with the RFN element were added to rifD transcripts in the presence and absence of FMN, and the resulting mixtures were digested with RNase H (which specifically cleaves RNA:DNA heteroduplexes) and analyzed by PAGE (A. S. Mirnov, et al., Cell 2002, 111, 747). A significant portion of transcripts held certain oligonucleotides in the absence of FMN, but not in the presence of FMN, indicating that FMN stabilizes a structural rearrangement of rifD transcripts that in turn prevents annealing of the oligonucleotide.

b. Affinity and Specificity of the FMN-rifD Complex

If the RFN element serves as an aptamer for FMN, it should exhibit characteristics of a saturable receptor that has some ability to discriminate against related ligands. To obtain values for apparent dissociation constant (apparent Kd) for FMN, in-line probing assays were repeated with trace amounts of rifD RNA and increasing concentrations of FMN; the ligand concentration that correlates with half-maximal modulation of RNA structure should reflect the apparent Kd. These experiments indicate that the rifD RNA contains a saturable ligand-binding site that exhibits an apparent Kd of ~5 mM. Furthermore, the RNA discriminates against the dephosphorylated form of FMN (riboflavin) by approxi-
mately three orders of magnitude. This exceptional ligand specificity of the ribD mRNA is surprising since the aptamer must generate a binding pocket for FMN that makes productive interactions with a phosphate group.

**[0350]** ii. FMN-Induced Transcription Termination

**[0351]** a. In Vitro Transcription Termination Mediated by an FMN Riboswitch

**[0352]** The relative amounts of the major transcription products for the ribD leader region were examined by in vitro transcription using T7 RNA polymerase or *Bacillus subtilis* RNA polymerase. The ribD leader region contains a classical intrinsic terminator just upstream of the ribD coding region. Interestingly, transcripts that terminated at the intrinsic terminator are specifically induced by FMN, in the absence of additional protein factors. Furthermore, mutations in the RFN element abrogate this phenomenon. The left-half of the terminator sequence forms alternative base-pairing interactions with a portion of the RFN element, thereby forming an antiterminator element. Sequence alterations of the intrinsic terminator eliminate FMN-induced termination while alterations in the antiterminator result in constitutive termination. Taken together, these observations are consistent with a mechanistic model wherein FMN directly interacts with ribD transcripts during conditions of excess FMN. Complex formation subsequently induces transcription termination within the 5'UTR (FIG. 12), which precludes gene expression by preventing the ORF from being transcribed. During conditions of limiting FMN, an antiterminator structure is formed within the ribD nascent transcript, which allows for synthesis of the downstream genes.

**[0353]** b. FMN-Mediated Control of Transcription Termination In Vivo

**[0354]** The molecular details of riboswitch-mediated transcription termination are likely to be more complex than this rather simplistic model implies. For example, given that the ‘decision’ to form the terminator or antiterminator conformation occurs only once during transcription, the regulatory mechanism is likely to rely on precise transcriptional kinetics as well as the appropriate RNA folding pathways. Moreover, the kinetics of FMN interacting with the RNA receptor is likely a critical factor. Although the affinity that the RNA has for FMN is exceptionally strong compared to engineered aptamers, it is possible that the kinetics of ligand association might be the more important determinant of genetic regulation. Indeed, all of these parameters are likely to conspire together in order to exert appropriate control over the intrinsic terminator. In adapting and designing riboswitches for use as described herein, the impact of transcription speed should be taken into account.

**[0355]** iii. Control of Transcription Termination by Other Riboswitches

**[0356]** Intrinsic terminators can be identified via computer-assisted search algorithms (F. Lillo, et al., 2002, 18, 971). Using such bioinformatic analyses, a subset of riboswitch RNAs that are predicted to contain an intrinsic terminator and an alternate antiterminator structural element can be identified (M. Mandal, et al., *Cell* 2003, 113, A. G. Vitreschkuk, et al., *Nucleic Acids Research* 2002, 30, 3141; F. J. Grundy, T. M. Henkin, *Molecular Microbiology* 1998, 30, 737; S. Kochhar, H. Paulus, *Microbiology* 1996, 142, 1635; D. A. Rodionov, et al., *Journal of Biological Chemistry* 2002, 277, 48949). Therefore, the results described above for the FMN riboswitch are indicative of the mechanisms used by many other riboswitch RNAs. Indeed, SAM- and TFP-dependent riboswitches have been demonstrated to exert control over transcription via formation of mutually exclusive intrinsic terminator and antiterminator structures (see, e.g., Example 7). Furthermore, mutations that disrupt and subsequently restore helices within the SAM riboswitch aptamer result in loss and restoration, respectively, of SAM binding. Concurrently, these mutations also result in disruption or restoration of SAM-induced transcription termination in accordance with ligand-binding function. Riboswitches can be adapted and designed to exert control over transcription termination signals that differ appreciably from classical intrinsic terminators according to principles described herein. As described elsewhere herein, expression platform domains having expression-controlling stem structures can be matched to aptamer domains by designing the PI stem of the aptamer domain such that the control strand (P1b) of the aptamer can form a stem structure with the regulated strand (P1c) of the expression platform.

**[0357]** 4. Riboswitch Regulation of Translation Initiation in Bacteria

**[0358]** An alternative mechanism of genetic control by riboswitches is the modulation of translation initiation. Unlike transcription termination, the entire mRNA would be synthesized by RNA polymerase, but expression would be prevented by the riboswitch until the metabolite concentration reached an acertain level. In most instances, it was observed that riboswitches prevent translation initiation in the presence of high concentrations of target metabolite. However, riboswitches can be designed and adapted such that allosteric modulation of riboswitch structures could lead to translation activation. The regulatory mechanism of translation control is briefly described below for a TPP riboswitch from *E. coli*.

**[0359]** i. A Natural Aptamer for TPP

**[0360]** A conserved RNA element, referred to as the thi box, was identified within 5'UTRs of mRNAs that are responsible for thiamine biosynthesis and transport (D. A. Rodionov, et al., *Journal of Biological Chemistry* 2002, 277, 48949; J. Miranda-Rios, M. Navarro, M. Soberon, *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98, 9736). Genetic experiments confirmed that this structural element was required for thiamine-dependent regulation of *Rhizobium meliloti* thiamine biosynthesis genes (J. Miranda-Rios, M. Navarro, M. Soberon, *Proceedings of the National Academy of Sciences of the United States of America* 2001, 98, 9736), yet no regulatory factor had been identified through classical genetic experimentation. Therefore, it was possible that the thi box might serve as a portion of a riboswitch that responds to thiamine or its derivatives.

**[0361]** In *E. coli*, thiamine biosynthesis and transport genes are primarily located within three operons and four single genes (T. P. Begley, et al., *Archives of Microbiology* 1999, 171, 293), wherein each operon is preceded by a thi element. To begin to assess the regulatory properties of these sequences, the leader regions for the thiMD and thiCFSHl operons were utilized to construct transcriptional and translational fusions to a lacZ reporter gene (see Example 2). Addition of exogenous thiamine results in repression of the lacZ reporter gene in *E. coli*. Results from these data demonstrate that the thiM gene is regulated primarily at the level of translation while the thiC leader region confers both transcriptional and translational regulation to the lacZ reporter.
[0362] a. Direct Binding of Thiamine Pyrophosphate by E. coli mRNAs

[0363] As described above for the FMN aptamer, direct binding of TPP to the thIM and thIC leaders was demonstrated by in-line probing assays (see Example 2). The addition of thiamine, thiamine monophosphate (TP), or the pyrophosphate derivative (TPP) leads to structural rearrangement of the thIM RNA, particularly in the region encompassing the thi element (Fig. 13). Significantly, TPP, which is the biocatalytic form of thiamine, exhibits the best affinity between the ligands, with an apparent $K_d$ of 500 nM, while TP and thiamine associate to thIM with apparent $K_d$ values of 3 µM and 40 µM, respectively. In-line probing assays of RNAs resembling the thIC leader region reveal even more dramatic discrimination between thiamine and its phosphorylated forms, exhibiting greater than a 1,000-fold difference between binding of thiamine and TPP. These data are consistent with genetic experiments that suggested that TPP synthesis was required for regulation (E. Webb, et al., Journal of Bacteriology 1996, 178, 253; E. Webb, D. Downs, Journal of Biological Chemistry 1997, 272, 15702). Also, this system provides another example of a natural RNA aptamer that makes productive contacts to phosphate groups.

[0364] b. Confirmation of TPP Binding by Equilibrium Dialysis

[0365] RNAs resembling the thIM leader region were synthesized and placed into one of a two-channel equilibrium dialysis apparatus, in which the compartments are separated by a 3,000-dalton molecular-weight-cut-off dialysis membrane. $^3$H-Thiamine was preferentially retained within the thIM-containing chamber when allowed to equilibrate between chambers (see Example 2). This effect could be eliminated by providing excess unlabeled thiamine, but could not be reversed when supplemented with oxothiamine, a close chemical analog of thiamine. Additionally, a mutated version of thIM was unable to shift $^3$H-thiamine to the RNA-containing chamber. Together, these data are indicative of the formation of stable thIM:phosphate complexes, wherein the sequence of the RNA and the chemical form of the ligand are critical for maximal binding affinity.

[0366] ii. Binding of Thiamine Derivatives Correlates with Structural Modulation

[0367] Close inspection of in-line probing data for thIM reveal two surprising patterns of structural modulation. First, the relative rates of spontaneous fragmentation between reactions containing either thiamine or TPP differ within an internal loop of the thi element (Fig. 13). Nucleotides in this region adopt an increase in structural order in the presence of TPP but not with thiamine, implying this region is somehow involved in formation of a pyrophosphate-recognition pocket. Secondly, the region of the SD sequence is the only portion outside of the thi element that becomes structurally modulated in the presence of TPP.

[0368] Specifically, the SD sequence exhibits a significant decrease in spontaneous cleavage relative to reactions lacking TPP, suggesting that the SD is converted into a more structurally constrained form upon binding of TPP. This idea is consistent with a mechanism (Fig. 13) whereby in the absence of TPP the SD has a significant degree of single-stranded character and is accessible for translation initiation. An anti-SD sequence is proposed to interact with an anti-anti-SD sequence within the TPP aptamer under these conditions. In contrast, during conditions of excess TPP, a TPP-RNA complex is formed that disrupts the base pairing of the anti-SD sequence, which is then free to interact directly with the SD and decrease the single-stranded character of the region, hence decreasing efficiency of translation initiation. Preliminary site-directed mutagenesis of the thIM mRNA supports this overall model (see Example 2). Specifically, mutations that disrupt TPP binding also disrupt regulation of translation for thIM-lacZ fusions, while mutations that alter the anti-SD sequence affect regulation but do not affect TPP binding. Thus, binding of thiamine correlates with both the structural accessibility of the SD and the translation efficiency in vivo.

[0369] iii. Control of Translation Initiation by Other Riboswitches

[0370] Bioinformatics analyses are consistent with molecular mechanisms similar to that of thIM also being recurrent amongst riboswitch RNAs. Specifically, anti-SD and anti-anti-SD structures have been proposed for several riboswitch classes, including FMN (A. G. Vitreschak, et al., Nucleic Acids Research 2002, 30, 3141), lysine, TPP (D. A. Rodionov, et al., Journal of Biological Chemistry 2002, 277, 48949), coenzyme H$_2$ (see Example 1) and SAM. In general, riboswitches from Gram-negative organisms seem to favor expression platforms that exert control over translation, while riboswitches from Gram-positive bacteria appear to predominate use expression platforms that control transcription termination. The latter can reflect a greater reliance upon multi-gene transcriptional units in Gram-positive organisms, which might be more efficient to preclude transcription of long operons when the gene products are unnecessary.

[0371] Biochemical evidence for riboswitch-mediated control over translation initiation has also been obtained for FMN and AdoCbl riboswitches (see Example 1). FMN binding to a riboswitch that regulates the B. subtilis ypaA gene results in alteration of the SD structural context, similar to what was observed for thIM. Interestingly, this genetic control element has also been proposed to regulate ypaA transcription (J. M. Lee, et al., Journal of Bacteriology 2001, 183, 7371), although the leader region does not contain an obvious intrinsic terminator structure. Binding of AdoCbl to the E. coli buiB riboswitch has also been demonstrated to correlate with regulation of translation in vivo.

[0372] Certain riboswitch RNAs exert control over transcription and translation using the same RNA sequence. For this class of riboswitches, the SD sequence is contained within an intrinsic terminator. Therefore, the formation of the terminator structure also enacts formation of a SD-sequestering structure. In total, all of these observations suggest that although the thIM and rib3 riboswitches represent useful paradigms for riboswitch-mediated control of translation and transcription, respectively, there are likely to be a wide variety of molecular mechanisms utilized by riboswitch RNAs for control of gene expression. Indeed, TPP riboswitches that must be employing different mechanisms of control have been identified in several plant and fungal species (see Example 4). The placement of these RNAs near splice sites in some instances and in the 3'UTR in others indicate TPP-responsive control over splicing and mRNA stability or expression, respectively.

[0373] 5. Early Origins?

and adenine riboswitches are also represented in numerous different genera, although they appear to be primarily limited to Gram-positive bacteria, with a few Gram-negative bacteria as exceptions (see Example 6). In all instances, the structural and sequence conservation of riboswitch classes is limited to the aptamer domain (FIG. 11). This is not unexpected given that the aptamer RNA must preserve its capability to bind the target chemical, which has not been significantly modified through evolution. In contrast, there is considerable sequence and structural diversity between expression platforms, even between riboswitches of the same class and within the same organism. Together, these data hint that the ligand-binding properties of riboswitch aptamer domains have been maintained throughout expansive evolutionary timescales.

TABLE 1-continued

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Transcriptional Unit</th>
<th>Predicted Gene Function(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>yuaJ</td>
<td>Unknown; putative thiamine transporter</td>
<td></td>
</tr>
<tr>
<td>yluB</td>
<td>Similar to acetylornithine decarboxylase</td>
<td></td>
</tr>
<tr>
<td>Guanine</td>
<td>yxaA</td>
<td>Similar to pyrimidine nucleoside transport</td>
</tr>
<tr>
<td></td>
<td>xpt-phuX</td>
<td>Xanthine permease</td>
</tr>
<tr>
<td></td>
<td>pheG</td>
<td>Hypoxanthine/ Guanine permease</td>
</tr>
<tr>
<td></td>
<td>purE-purK-purB-purC- purG-purQ-purG-purF- purM-purN-purH-purD</td>
<td>Purine biosynthesis</td>
</tr>
<tr>
<td>Adenine</td>
<td>ydhL</td>
<td>Unknown</td>
</tr>
<tr>
<td>S-adenosylmethionine</td>
<td>ytf</td>
<td>Putative myeine tetrahydrofolate reductase</td>
</tr>
<tr>
<td></td>
<td>metI-metC</td>
<td>Methionine biosynthesis</td>
</tr>
<tr>
<td></td>
<td>ykr-TykrS</td>
<td>5’ methylthioadenosine recycling pathway</td>
</tr>
<tr>
<td></td>
<td>ykrW-ykrX-ykrY-ykrZ</td>
<td>5’ methylthioadenosine recycling pathway</td>
</tr>
<tr>
<td></td>
<td>cysI-cysD-cysE-cysC- cysD-yiaF-yiaF</td>
<td>Cysteine biosynthesis</td>
</tr>
<tr>
<td></td>
<td>yoaD-ycdC-yoaB</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>metE</td>
<td>Methionine synthase, B20- independent</td>
</tr>
<tr>
<td></td>
<td>metK</td>
<td>S-adenosylmethionine synthetase</td>
</tr>
<tr>
<td></td>
<td>yusC-yusB-yusA</td>
<td>Unknown ABC transporter</td>
</tr>
<tr>
<td></td>
<td>yxG</td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>yxH</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Gene nomenclature is derived from the SubList database except for metE and metC, which are recent designations (S. Anger, et al., Microbiology 2002, 148, 973).


There is clear precedence for the targeting of RNAs with small molecule drugs (G. J. Zaman, et al., Nucleic Acids Research 2002, 30, 62), the most obvious example being that of ribosomal RNA. Several other bacterial-specific RNAs have been explored as candidates for small molecule drug interaction; however, the approach relies upon screening large chemical libraries for those chemicals that fortuitously interact with the RNA of interest, even though the RNA itself does not naturally form a binding pocket for small organic molecules. Riboswitch RNAs therefore exhibit an advantage in antimicrobial development given that they serve as a receptor for small molecule ligands, much like their protein receptor counterparts.

In addition to their use as targets for chemical inhibition, understanding of the mechanisms utilized by natural riboswitch RNAs allows adaptation of riboswitches and development of new riboswitches as novel genetic control elements. Numerous aptamer RNA sequences have been identified that interact with a wide variety of small organic molecules (M. Famulok, Current Opinion in Structural Biology 1999, 9, 324). Engineered riboswitches can be generated that respond to non-biological, or otherwise metabolically inert, compounds. Such genetic control elements can be used for a variety of expression control and molecular detection applications.
D. Example 4

Eukaryotic Riboswitches

[0380] 1. Abstract

[0381] Genetic control by metabolite-binding mRNAs is wide spread in prokaryotes. These "riboswitches" are typically located in non-coding regions of mRNA, where they selectively bind their target compound and subsequently modulate gene expression. Disclosed are mRNA elements that have been identified in fungi and in plants that match the consensus sequence and structure of thiamine pyrophosphate-binding domains of prokaryotes. In Arabidopsis, the consensus motif resides in the 3'-UTR of a thiamine biosynthetic gene, and the isolated RNA domain binds the corresponding coenzyme in vitro. These results suggest that metabolite-binding mRNAs possibly are involved in eukaryotic gene regulation and that some riboswitches might be representatives of an ancient form of genetic control.

[0382] 2. Introduction

[0383] Riboswitches are genetic control elements that can be found in the 5'-untranslated region of certain messenger RNAs of prokaryotes (see Examples 1-3). These genetic switches exhibit two surprising properties. First, the mRNA is able to form a highly selective binding site for the target metabolite without the aid of proteins. Second, metabolite binding brings about an allosteric reorganization of RNA structure that leads to alterations in genetic expression. Unlike many other genetic control systems, riboswitches do not require metabolite-binding proteins to serve as sensors, and thus offer a direct link between the genetic information that is encoded by an mRNA and its chemical surroundings.

[0384] A number of distinct types of riboswitches have been confirmed by biochemical and genetic analyses. For example, a coenzyme B12-binding RNA has been shown (Example 1) to control expression of the Escherichia coli bteB gene, which encodes a cobalamin transport protein. Riboswitches triggered by thiamine pyrophosphate (TPP) have been shown to control operons in E. coli (Example 3) and Bacillus subtilis (Example 6) that are responsible for biosynthesis of this coenzyme. In addition, the RNF element, which frequently is found in the 5'-untranslated region of genes responsible for the biosynthesis or import of riboflavin and FMN, serves as the receptor portion of FMN-dependent riboswitches in Bacillus subtilis (see Examples 3 and 6). Recently, it has been determined that certain 5'-UTR motifs that are located in the 5'-UTRs of numerous genes in B. subtilis bind the coenzyme 5'-adenosylmethionine (SAM) with high affinity and precision. These findings indicate that riboswitches are used to recognize a diverse collection of metabolites and that direct sensing of small molecules by mRNAs is an important form of genetic control for certain organisms. Disclosed herein, is evidence that metabolite-binding domains are embedded in certain mRNAs of eukaryotes, indicating that higher organisms might also exploit riboswitches for genetic control.

[0385] 3. Results

[0386] Disclosed are many RNA elements that have been identified in prokaryotes that exhibit sequence similarity to the B12- and SAM-dependent riboswitches. Given the relatively large size and sequence complexity of these RNA motifs, it is unlikely that numerous evolutionary reinventions of the same elements would have occurred. Furthermore, the metabolite triggers of these genetic switches are predicted to have been present in a time before the emergence of proteins (White, 1976; Benner et al., 1989; Jeffares et al., 1998). This is consistent with the known classes of metabolite-sensing RNAs having originated in the ancient RNA world, which is believed to be a time before the emergence of proteins and when metabolism was guided entirely by RNA (Joyce, 2002).

[0387] If the present-day riboswitches are of ancient origin, then eukaryotes might possess RNA genetic switches that are descended from the last common ancestor of modern cells. Disclosed herein several eukaryotes carry RNA domains that conform to the consensus sequence and structure of the metabolite-binding domain of the TPP riboswitch class (FIG. 14A). (The mRNAs that carry the TPP-binding domains encode for a protein that is homologous to the thiC protein of E. coli. This protein enzyme catalyzes the conversion of 5-aminomimidazole ribotide (AIR) to hydroxymethylpyrimidine phosphate (HMP-P), which is a key biosynthetic step in the synthesis of thiamine and ultimately TPP (Vander Horn et al., 1993; Begley et al., 1999)). For example, a putative thiamine biosynthesis gene of Arabidopsis thaliana carries an RNA element (FIG. 14B) in its 3'-UTR that conforms to the consensus TPP-binding domain. Similar RNA elements are found in rice (Oryza sativa) and bluegrass (Poa secunda). RNA elements that conform to the TPP-binding sequence and structure are also present in fungi such as Neurospora crassa (FIG. 14C) and Fusarium oxysporum. As with plants, the riboswitch homologs in fungi are located in genes that have been implicated in the biosynthesis of thiamine, suggesting that in each case their role is to maintain required coenzyme levels by modulating expression of the appropriate biosynthetic genes. A sequence alignment of the homologous domains found in eukaryotes compared to that of the gram negative bacterium E. coli (thiC and thiM) and the gram positive bacterium Clostridium acetobutylicum (thiC) is depicted in FIG. 15.

[0388] The RNA element corresponding to the consensus TPP-binding domain of A. thaliana (FIG. 14A) was generated by in vitro transcription of a synthetic DNA template and the RNA was subjected to "in-line probing" (FIG. 16A). This method relies on the spontaneous breakdown of RNA phosphodiester linkages, whose pattern of cleavage can be used to reveal the structural and functional features of ligand-binding RNAs (see Examples 1-3). Indeed, the riboswitch-like element exhibits TPP-dependent structural modulation and has a fragmentation pattern that is consistent with the predicted secondary structure of TPP riboswitches from bacteria (see Examples 2 and 3). In addition, this structure-stabilizing method has been used herein to establish that the RNA binds TPP with an apparent dissociation constant (Kd) of ~50 nM (FIG. 16B), which is similar to that determined previously for an E. coli riboswitch variant. Similarly, it has been demonstrated that the sequence elements of fungi that correspond to the TPP riboswitch consensus also bind TPP with high affinity.

[0389] Sequestering of the ribosome binding site and transcription termination are demonstrated mechanisms for TPP riboswitches in E. coli (FIG. 17). Since the TPP-binding element in plants is located immediately upstream from the polyA tail, it is possible that metabolite binding might regulate mRNA processing and stability. Alternatively, a consensus TPP-binding sequence (FIG. 14C) identified in the fungal genome of N. crassa resides in an intron, suggesting that RNA splicing might also be guided by metabolite-binding pre-mRNAs. In prokaryotes, ligand binding typically brings about allosteric changes in the Watson-Crick base pairing arrangements near gene control elements such as transcr-
tion terminators and ribosome binding sites. Likewise, secondary structure rearrangements by metabolite-binding riboswitches can be used to modulate a greater variety of RNA processing, transport and expression pathways in eukaryotes.

[0390] Although it is likely that TPP-binding domains and those for coenzyme B₁₂, FMN, and SAM are of ancient origin, it is possible that other examples of metabolite-binding mRNAs have emerged more recently in evolution. These newer riboswitches would be more narrowly distributed across the phylogenetic landscape, so efforts to search for new riboswitches that are triggered by compounds that are not ancient and universally distributed will be difficult. Regardless of the scope of riboswitch use in modern organisms, both natural and engineered riboswitches could have significant utility. Given the central role that known riboswitches serve in modulating the concentration of key coenzymes, these RNAs can serve as new targets for drug discovery efforts. Therefore, reverse engineering of natural riboswitches can be used to establish a conceptual basis for creating designer riboswitches for the purposeful control of eukaryotic genes.

E. Example 5
Lysine Riboswitches

[0391] The precise control of gene expression in response to changes in the chemical and physical environment of cells requires selective interactions between biochemical sensor elements and the molecules that carry or interpret genetic information. Most known genetic factors that respond to such environmental changes are proteins (Pashine and Gunn 2002). However, a number of studies (e.g. see Examples 1-3 and 6-8) have demonstrated that natural RNA molecules can also recognize small organic compounds and harness allosteric changes to control the expression of adjacent genes. These metabolite-binding RNA domains, termed riboswitches, typically are embedded within the 5’-UTRs of mRNAs and control the expression of proteins involved in the biosynthesis or import of the target compound. Riboswitches also play an important role in controlling fundamental metabolic pathways in bacteria involved in sulfur metabolism, and in the biosynthesis of various coenzymes and purines (see Example 6). Furthermore, riboswitches are phylogenetically widespread amongst eubacterial organisms, and both sequence and biochemical data suggest that riboswitches are also present in the genes of eukaryotes (see Example 4).


[0393] 1. Materials and Methods

[0394] i. Chemicals and Oligonucleotides

[0395] L-lysine, all analogs with the exception of L-α-homolysine (compound 6, FIG. 20A), tritiated lysine (L-Lysine-[4,5-3H(N)]), and the four dipetides were purchased from Sigma. A protocol adapted from that reported previously (Dong, Z. 1992, Tetrahedron Lett. 33:7725-7726) was used to synthesize L-α-homolysine. Purity and integrity of synthetic L-α-homolysine was confirmed by TLC and NMR.

[0396] DNA oligonucleotides were synthesized by the HHMI Keck Foundation Biotechnology Resource Center at Yale University, purified by denaturing PAGE and eluted from the gel by crush-soaking in 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl, and 1 mM EDTA. Oligonucleotides were recovered from solution by precipitation with ethanol.

[0397] ii. Phylogenetic Analyses

[0398] L box domains were identified by sequence similarity to the B. subtilis lysC 5’-UTR. Ultimately, the program was used to search for degenerate matches to the pattern (WAGAGNGNC [10] A [3] RTA [50] RRGR [10] CCCGAR [40] GG [13] VAA [13] YTGCTCA [36] TGRWG [2] CTYW) (SEQ ID NO:376), however, less complete versions of this pattern were used with iterative refinements to identify the consensus sequence and structure of the L box motif. Fracteted numbers are variable gaps with constrained maximum lengths denoted. Nucleotide notations are as follows: Y=pyrimidine; R=purine; W=A or T; K=G or T; V=A, G or C. Up to six violations of this pattern were permitted when forming the phylogeny depicted in FIG. 18.

[0399] iii. In-Line Probing of RNA Constructs

[0400] The B. subtilis 315 lysC, 237 lysC and 179 lysC RNAs were prepared by in vitro transcription using T7 RNA polymerase and the appropriate PCR DNA templates. RNA transcripts were dephosphorylated and subsequently 5’-32P-labeled using a protocol similar to that described previously (Seetharaman, S., et al., 2001, Nature Biotechnol. 19, 336-341). Labeled precursor RNAs (~2 nM) were subjected to in-line probing using conditions similar to those described in Examples 1 and 2. Reactions (10 µL) were incubated for 40 hr at 25°C in a buffer containing 50 mM Tris (pH 8.5 at 25°C), 20 mM MgCl₂, and 100 mM KCl in the presence or absence of L-lysine or various analogs as indicated for each experiment. Denaturing 10% PAGE was used to separate spontaneous cleavage products, which were detected and quantitated by using a Molecular Dynamics PhosphorImager and ImageQuaNT software.

[0401] iv. Equilibrium Dialysis and Scatchard Analyses

[0402] Equilibrium dialysis assays were conducted using a DispoDialysis Dialyzer (ED-1, Harvard Bioscience), wherein two chambers a and b were separated by a 5,000
MWCO membrane. The final composition of buffer included 50 mM Tris-HCl (pH 8.5 at 25°C), 20 mM MgCl₂ and 100 mM KCl (50 µL delivered to each chamber). Assays were initiated by the addition of [1-3H]-lysine (50 µM initial concentration prior to equilibration; 40 Ci mmol⁻¹; 15,000 cpm) to chamber a. When present, RNA (179 lysC) was introduced into chamber b to yield a concentration of 10 µM. After 10 hr of equilibration at 25°C, a 3-µL aliquot from each chamber was removed for quantification by liquid scintillation counter. Competition assays were established by delivering an additional 3 µL of buffer to a and an equivalent volume of buffer containing 50 µM unlabeled L-lysine, D-lysine, L-ornithine, or L-lysine hydroxamate as indicated to b. After 10 hr of additional incubation at 25°C, 3-µL aliquots were again drawn for quantification of tritium distribution.

[0403] Scatchard data points were generated as described above with the following exceptions. RNA was added to chamber b to yield a concentration of 1 µM RNA and equilibration of the dialysis mixtures proceeded for 20 hr. In addition, [3H]-lysine concentrations were varied from 50 nM to 2.5 µM. Calculation of points on the Scatchard plot from the equilibrium dialysis data was carried out as described elsewhere herein.

[0404] vi. In Vitro Transcription Termination Assays

[0405] Transcription termination assays were conducted using a method of single-round transcription adapted from that described previously (Landick, R., et al., 1986, Methods Enzymol. 274:334-353). The template for lysC 5'-UTR transcription was altered (CGG of the RNA) such that the first C residue of the nascent RNA is not encountered until position 17. Polymerization was initiated by the addition of a mixture of Apa dinucleotide (1.35 µM), GTP and UTP (2.5 µM each) plus unlabeled ATP (1 µM) and [α-32P]-ATP (4 µCi), which was incubated for 10 min. Halted complexes are restarted by the addition of 150 µM each of the four NTPs, and heparin (0.1 mg ml⁻¹) is simultaneously added to prevent polymerases from initiating transcription on new templates. Transcription mixtures also contained 20 mM Tris-HCl (pH 8.0 at 23°C), 20 mM NaCl, 14 mM MgCl₂, 0.1 mM EDTA, 0.01 mg/ml BSA, 1% w/v glycerol, 4 pmol DNA template, 0.045 U µL⁻¹ E. coli RNA polymerase (Epicenter, Madison, Wis.), and 10 mM of L-lysine or the lysine analog as indicated for each experiment. Reactions were incubated for an additional 20 min at 37°C, and the products were examined by denaturing 6% PAGE followed by analysis using a PhosphorImager.

[0406] vii. In Vivo Analysis of lysC Genetic Variants

[0407] Fusions of the lysC 5'-UTR with a lacZ reporter gene were used to assess the function of the lysine riboswitch in vivo using methods similar to those described elsewhere herein. Briefly, the lysC 5'-UTR, comprising the promoter and the first 315 nucleotides of the transcription template, was prepared as an EcoRI-BamHI fragment by PCR. Sequence variants M1 through M3, G39A, and G40A were generated by PCR amplification of the wild-type construct using primers that carried the desired mutations. The PCR products were cloned into pBG1661 immediately upstream of the lacZ reporter gene and the integrity of the resulting clones were confirmed by sequencing. Transformations of pBG1661 variants into B. subtilis strain 1A40 (obtained from the Bacillus Genetic Stock Center, Columbus, Ohio) were performed and the correct transformants were identified by selecting for chloramphenicol resistance and screening for spectinomycin sensitivity.

[0408] Cells were grown with shaking at 37°C. either in rich medium (2XYT broth or tryptophosphate blood agar base) or defined medium (0.5% w/v glucose, 2 g L⁻¹ (NH₄)₂SO₄, 18.3 g L⁻¹ K₂HPO₄, 3H₂O, 6 g L⁻¹ KH₂PO₄, 1 g L⁻¹ sodium citrate, 0.2 g L⁻¹ MgSO₄·7H₂O, 5 µM MnCl₂, and 5 µM CaCl₂). Methionine, lysine, and tryptophan were added to 50 µg ml⁻¹ for routine growth. Growth under lysine-limiting conditions was established by incubation under routine growth conditions in defined medium to an A₆₀₀ of 0.1, at which time the cells were pelleted by centrifugation, resuspended in minimal medium, split into five aliquots, and supplemented with five different media types as defined in the legend to FIG. 22C. Cultures were incubated for an additional 3 hr before performing β-galactosidase assays.

[0409] 2. Results

[0410] i. The L Box: A Conserved mRNA Element that is Important for Genetic Control

[0411] Riboswitches are typically formed by close juxtaposition of a metabolite-binding ‘aptamer’ domain and an ‘expression platform’ that interfaces with mRNA elements necessary for gene expression. Although the RNA sequences and structural components that serve as the expression platform change significantly throughout evolution, the aptamer domain largely retains the sequence composition of its ligand-binding core along with the major secondary-structure features. This permits the use of phylogenetic analyses to identify related RNA domains and to establish a consensus sequence and structure for a given class of riboswitches.

[0412] Beginning with the sequence homology reported to exist between the lysC 5'-UTRs of three bacterial species (Pate, J.-C., et al., 1998, FEMS Microbiol. Lett. 169:165-170), the number of representatives was expanded using an algorithm that searches for related sequences and secondary structures (e.g. see Examples 4 and 6). 31 representatives of this RNA domain, termed the “L box”, in the 5'-UTRs of lysC homologs and other genes related to lysine biosynthesis from a number of Gram-positive and Gram-negative organisms were identified (FIG. 18). The sequence alignment reveals that the RNA forms a five-stem junction wherein major base-paired domains are interspersed with 56 highly conserved nucleotides (FIG. 19A). Furthermore, the base-paired elements P2, P2a, P2b, P3 and P4 each appear to conform to specific length restrictions, suggesting that they are integral participants in the formation of a highly structured RNA. It was also noticed that conserved sequences in the junction between stems P2 and P2a conform to a “loop L” motif, which is an RNA element that occurs frequently in other highly-structured RNAs (e.g. see Leonitis, N. B., and Westhof, E. 1998, J. Mol. Biol. 283:571-583).

[0413] The L box domain of the B. subtilis lysC mRNA resides immediately upstream from a putative transcription terminator stem (Kochhar, S., and Paulus, H. 1996, Microbiol. 142:1635-1639; Patte, J.-C., et al., 1998, FEMS Microbiol. Lett. 169:165-170). In several other riboswitches with similar arrangements (e.g. Examples 3 and 6), the 5'-UTR can be trimmed to separate the minimal aptamer domain from the adjacent expression platform. An RNA fragment (237 lysC, FIG. 19B), encompassing nucleotides 1 through 237 of the lysC 5'-UTR, was generated and examined for allosteric function. This construct, which excludes the putative transcription terminator stem, was subjected to structural analysis by inline probing (Soukup, G. A. and Breaker, R. R. 1999, RNA 5:1308-1325) to determine whether the presence of lysine alters RNA structure. It was observed that 237 lysC exhibits a
pattern of spontaneous RNA cleavage (FIG. 19C) that is consistent with the secondary structure model of the L box motif constructed from phylogenetic sequence data. Furthermore, it was found that the addition of 10 μM L-lysine causes significant changes in the cleavage pattern at four locations along the RNA chain, indicating that allosteric modulation of the 5'-UTR fragment is occurring. In addition, the same pattern of spontaneous cleavage and amino acid-dependent structural modulation was observed when using the 179 lysC RNA construct, which encompasses only the most highly-conserved portion of the L-box motif (nucleotides 27 through 205 of the lysC 5'-UTR).

[0414] A reduction of spontaneous cleavage is observed in each of the four sites of metabolite-induced structural modulation. In most instances, a reduction in spontaneous cleavage is due to the nucleotides becoming more ordered in the complex formed between RNA and its ligand (Soukup, G. A. and Breaker, R. R. 1999, RNA 5:1308-1325). Interestingly, these four groups of nucleotides are located at the center of the 5'-stem junction of the L box secondary structure model (FIG. 19B), implying that these nucleotides are directly involved in recognizing the amino acid target. Similar patterns of ligand-induced structural modulation have been observed with the aptamer domains of other riboswitches (see Examples 2, 3 and 6).

[0415] ii. The Lysine Aptamer Exhibits High Specificity for L-Lysine and Discriminates Against Closely-Related Analogs

[0416] Riboswitches, like their counterpart genetic factors made of protein, must exhibit sufficient specificity and affinity for their target metabolite in order to achieve precision genetic control. To examine the molecular recognition characteristics of the lysC L box domain, a series of in-line probing assays were performed using various analogs of lysine at 100 μM. The properties of a lysine analog collection were examined, wherein each compound carries minimal chemical changes relative to L-lysine (FIG. 20A). Nearly every chemical alteration to the amino acid renders the compound incapable of causing a structural modulation of the 179 lysC RNA (FIG. 20B). Perhaps most striking is that the RNA does not undergo structural modulation in the presence of D-lysine, which differs from L-lysine by the stereochemical configuration at a single carbon center.

[0417] The absence of significant structural modulation in the presence of D-lysine and of other analogs indicates that at least three points of contact are being made between the RNA and its amino acid target. Specifically, the observation that analogs 1, 3, and 4 fail to induce structural modulation is consistent with contacts being made to the amino and carboxy groups of the chain atoms, and to the amino group of the side chain, respectively. Moreover, the failures of compounds 2, 5, 6, 7 and 8 to induce conformational change in the RNA indicate that the aptamer forms a highly discriminating binding pocket that can measure the length and the integrity of the alkyl side chain. This high level of molecular discrimination is of particular biological significance, as a genetic switch for lysine most likely must respond exclusively to L-lysine and not closely related natural compounds.

[0418] Similarly, the allosteric response of the 179 lysC RNA to various dipeptides and acid-hydrolyzed dipeptides was examined. It was hypothesized that dipeptides should not trigger allosteric modulation of RNA structure, but that acid-mediated hydrolysis of dipeptides (FIG. 20C) carrying at least 1 lysyl residue should become active. As predicted, 179 lysC does not undergo allosteric modulation upon the addition of the dipeptides lys-lys, lys-ala, ala-lys, or ala-ala (FIG. 20D). However, the three dipeptides that carry at least one lysyl residue induce structural modulation of RNA upon pre-treatment of the dipeptides with 6 N HCl at 115° C. for 23 hr, followed by evaporation and neutralization. The extent of structural modulation (FIG. 20E) indicates that the samples containing the hydrolyzed lysine-containing dipeptides fully saturate the lysC aptamer, which is in accordance with the acid-mediated release of saturating amounts (greater than 1 μM; see below) of L-lysine.

[0419] It was also observed that an intermediate level of structural modulation occurs when D-lysine is pre-treated with HCl. Interestingly, the published rate of epimerization between D- and L-lysine (Engel, M. H., and Hare, P. E. 1982. Racemization rates of the basic amino acids. Year Book Carnegie Inst. Washington 81:422-425) is sufficient to account for the approximately 1 μM of L-lysine that is needed to produce half-maximal structural modulation (FIG. 20E). These results are consistent with lysine acting as the molecular ligand for the lysC aptamer, and that RNA conformational changes are not due to unknown contaminants of the commercial L-lysine preparation.

[0420] iii. Binding Affinity and Stoichiometry of the B. subtilis L-Lysine Aptamer

[0421] An approximation of the dissociation constant (K_d) was made by conducting in-line probing assays with 179 lysC using various concentrations of L-lysine (FIG. 21A). The sites of structural modulation exhibit progressively lower levels of spontaneous cleavage in response to increasing concentrations of ligand. A plot of the extent of RNA cleavage versus concentration of L-lysine (FIG. 21B) indicates that half-maximal structural modulation occurs when approximately 1 μM amino acid is present in the mixture, thus reflecting the apparent K_d of the 179 lysC for its target ligand.

[0422] The apparent K_d value for a longer construct that encompasses structural elements predicted to be involved in transcription termination exhibits a significantly poorer affinity for L-lysine. Specifically, an RNA construct encompassing nucleotides 1 through 315 of the lysC 5'-UTR was found by in-line probing to exhibit an apparent K_d of ~500 μM. Similar differences in ligand affinities for other riboswitches have been observed, wherein the minimized aptamer binds more tightly its cognate ligand compared to the same aptamer in the context of the complete riboswitch (aptamer plus the adjoining expression platform). This is most likely due to the presence of competing secondary or tertiary structures that might be important for the function of the riboswitch as a genetic control element, but that reduce ligand binding affinity by reducing pre-organization of the aptamer domain.

[0423] Equilibrium dialysis also was used to examine the affinity and specificity of the 179 lysC aptamer for its target (FIG. 21C). In the absence of RNA, tritiated L-lysine is expected to distribute equally between the two chambers (a and b) of an equilibrium dialysis apparatus. However, the addition of excess aptamer to one chamber of the system should shift the distribution of tritium towards this chamber as a result of complex formation. This asymmetric distribution of tritium is expected to be restored to unity by the addition of a large excess of unlabeled competitor ligand, which displaces the bulk of the tritiated lysine from the RNA. As expected, the fraction of tritiated L-lysine in chamber b of the equilibrium dialysis apparatus is ~0.5 in the absence of RNA (FIG. 21C) after a 10 hr incubation. This fraction is
altered to ~0.8 after incubation when a 200-fold excess of 179 lysC (10 μM) is added to chamber b, while this symmetric distribution of tritium is restored upon incubation for an additional 10 hours after the introduction of excess (50 μM) unlabeled L-lysine. Furthermore, D-lysine and L-ornithine do not restore equal distribution of tritium, which is consistent with their failure to modulate RNA structure as determined by in-line probing.

[0424] A Scatchard plot also was created by using data from a series of equilibrium dialysis experiments conducted under various concentrations of 3H-labeled L-lysine (FIG. 21D). The slope of the resulting line indicates that the 179 lysC RNA binds to L-lysine with an apparent K_0 of ~1 μM, which is consistent with that observed by using in-line probing. Furthermore, the x-intercept of the line occurs near an r value of 1, which demonstrates that the RNA forms a 1:1 complex with its ligand.

[0425] iv. The Lysine Aptamer and Adjacent Sequences Function as an Amino Acid-Dependent Riboswitch

[0426] With a number of riboswitches examined to date, there is a discernible set of structures residing immediately downstream of the aptamer domain that serve to control gene expression in response to ligand binding. Typically, the structure of this “expression platform” is modulated by metabolite binding to the aptamer domain. The alternative structure subsequently leads to modulation of transcription or translation processes. For example, the TPP riboswitch on the thMM mRNA of E. coli carries an expression platform that appears to preclude ribosome binding to the Shine-Dalgarno sequence of the adjacent coding region (see Example 2). Similarly, the expression platforms of various riboswitches from B. subtilis undergo ligand-induced formation of a stem-loop structure that induces transcription termination (e.g. Examples 3, 6 and 7).

[0427] It has been reported that the lysC mRNA undergoes transcription termination in cultured B. subtilis cells grown in the presence of excess L-lysine (Kochhar, S., and Paulus, H. 1996, Microbiol. 142:1635-1639.). It was observed herein that a sequence domain that participates in forming the P1 stem of the lysC aptamer is complementary to a portion of the putative terminator hairpin that resides ~30 nucleotides downstream (FIG. 22A). This architecture is similar to that of several other riboswitches, some of which exhibit termination of transcription in vitro upon addition of the corresponding ligand as cited above. Therefore, the lysC leader sequence appears to serve as a L-lysine-specific riboswitch that induces transcription termination by modulating the formation of a terminator stem.

[0428] In vitro transcription assays were conducted in the absence and presence of L-lysine and several analogs (FIG. 22B, left). In the absence of added ligand, single-round transcription in vitro using E. coli RNA polymerase produces terminated product corresponding to ~36% of the total transcription yield. In contrast, the amount of terminated product increases to ~76% when 10 nM L-lysine is present during in vitro transcription. Neither D-lysine nor L-ornithine induce termination, which is consistent with the fact that these compounds are not recognized by the lysine aptamer domain and thus are not expected to trigger transcription termination.

[0429] The configuration of the expression platform for the lysC gene in B. subtilis strongly implies a transcription termination mechanism, wherein the binding of L-lysine is expected to stabilize the P1 stem, thus permitting formation of the terminator hairpin (FIG. 22A). This proposed mechanism was examined by placing mutations within the critical pairing elements and by assessing lysine-induced transcription termination (FIG. 22B, center). Specifically, variant M1 carries two mutations that disrupt the formation of the terminator stem. This variant loses lysine-dependent modulation of transcription termination, and produces greater transcriptional read-through relative to the wild-type construct. M2 carries a total of four mutations that compensate for the disruption of the terminator stem, but that cause disruption of the anti-terminator stem. This construct also loses lysine-dependent modulation, whereas the amount of the terminated product expectedly becomes greater. Finally, the six-nucleotide variant M3 that carries the same mutations as M2 plus two additional mutations to restore the anti-terminator base-pairing potential results in near wild-type performance with regards to lysine-mediated modulation of transcription termination. These findings are consistent with a riboswitch mechanism wherein lysine binding precludes formation of an anti-terminator stem, thus increasing transcription termination by formation of an intrinsic terminator structure.

[0430] v. Evidence that Riboswitches Serve as Antibiotics Targets

[0431] Unlike other lysine analogs, both L-lysine hydroxymate and the antimicrobial compound thiosine (S-(2-aminoethyl)-L-cysteine; FIG. 22A, inset) cause an increase in transcription termination (FIG. 22B, left). These two compounds exhibit the best apparent K_0 values of any of the analogous tested, with values for both L-lysine hydroxymate and thiosine of ~100 μM and ~30 μM, respectively (data not shown). In previous studies, a series of mutants were identified in B. subtilis (Vold, B., et al., 1975, J. Bacteriol. 121:970-974; Lu, Y., et al., 1992, FEMS Microbiol. Lett. 92:23-27) and E. coli (Pacca, J.-C., et al., 1998, FEMS Microbiol. Lett. 169:165-170) that cause resistance to thiosine and cause derepression of lysC expression. These mutations all map to the lysine aptamer domain (see FIG. 22A for select B. subtilis mutants), and all appear to cause disruptions in the conserved elements or the base-pairing integrity of the structure.

[0432] The functional integrity of two thiosine-resistant mutants (G39A and G40A) was examined by equilibrium dialysis and by in line probing, and both mutants fail to exhibit lysine-binding activity. Furthermore, RNA constructs that carry mutations in the otherwise conserved P1-P2 junction fail to undergo lysine-dependent transcription termination in vitro (FIG. 22B, right). These findings suggest that the antimicrobial action of thiosine might at least partially be due to direct binding of the analog to the lysine riboswitch, causing repression of aspartokinase expression to a level that is deleterious to cell growth.

[0433] The function of the wild-type 5’-UTR of lysC and of the two thiosine-resistant mutants were also examined in vivo by fusion to a lacZ reporter gene. The wild-type riboswitch domain exhibits ligand-dependent modulation upon addition of L-lysine, whereas the G39A and G40A mutants fail to regulate β-galactosidase expression (FIG. 22C, medium II versus III). In contrast, lysine hydroxymate fails to repress expression of the reporter gene in vivo (medium IV), indicating that this compound might not attain a sufficiently high concentration inside cells to trigger transcription termination. As with lysine, thiosine also represses β-galactosidase expression for the wild-type construct, but not the two derepression mutants (medium V). This latter observation is consistent with the antimicrobial action of thiosine being due largely to its function as an effector for the lysine riboswitch.
3. Conclusions

The first mutants that caused deregulation of lysine biosynthesis in *B. subtilis* were identified nearly three decades ago (Vold, B., et al., 1975, *J. Bacteriol.* 121:970-974), however, the mechanism of genetic regulation has remained unresolved. Disclosed herein, it was demonstrated that the 5'-UTR of the lysC mRNA from *B. subtilis* serves as a riboswitch that responds to the amino acid lysine. The derepressed mutants isolated in the original study cause disruption of the aptamer domain of the riboswitch, such that the ligand is no longer bound by the RNA. Furthermore, in vivo expression studies using mutant lysC fragment-reporter gene fusions indicate that these riboswitch mutations most likely cause unregulated over-expression of aspartokinase, which catalyzes the first step in the biosynthetic pathway to lysine and several other amino acids.

Bacteria use various mechanisms to respond genetically to amino acid concentrations. Two of the more prominent mechanisms, translation-mediated transcription attenuation and T box-dependent mechanisms (Henkin, T. M., and Yanofsky, C. 2002, *BioEssays* 24:700-707), both sense the presence of non-aminoacylated tRNAs. Indeed, 18 of the 20 common amino acids in *B. subtilis* appear to be detected indirectly through the use of T box elements. Interestingly, there is no known tRNA<sup>ψ</sup>-dependent T-box in any organism, and presumably the lysine riboswitch described herein serves as the genetic sensor for this amino acid in the absence of a corresponding T box. Moreover, the genetic distribution of lysine riboswitches affiliated with the nhaC gene from several organisms indicates that this RNA genetic element might be a key regulator of cellular pH.

Since the lysC mRNA functions as a receptor for L-lysine, the Lys riboswitch can serve as a drug target. (See other examples, Hesselberth, J. R., and Ellington, A. D. 2002, *Nature Struct. Biol.* 9:891-893; Sudarsan, N., et al., 2003, *RNA* 9:644-647). The lysine riboswitch, and perhaps other classes of riboswitches as well, can be targeted by analogs that selectively bind to the riboswitch and induce genetic modulation. In *B. subtilis*, an analog of lysine that triggers the riboswitch would be expected to function as an antimicrobial agent, because the reduction of aspartokinase expression should induce starvation for lysine and other critical metabolites. The finding that thiosine binds to the lysine aptamer in vitro, and causes down regulation of a reporter construct fused to the wild-type riboswitch, provides support for the view that riboswitches are a newly recognized class of targets for drug discovery.

Recent discoveries have been elucidating the roles of small RNAs in guiding gene expression in a wide range of organisms (for a review see Gottesman, S. 2002, *Genes Dev.* 16:2829-2842). It is apparent that small RNAs, including riboswitch domains embedded within mRNAs, can control gene expression by a wide range of mechanisms. Unlike other RNA genetic control elements, riboswitches directly bind to metabolites and control the expression of genes that are involved in the import and biosynthesis of a number of fundamental metabolites. Riboswitches examined previously respond to compounds that are chemically related to nucleotides. However, the existence of a class of riboswitches that responds to a small amino acid with high selectivity serves as proof that natural RNA switches can detect and respond to a greater range of metabolic classes.

F. Example 6

Guanine and Other Riboswitches in *Bacillus subtilis* and Other Bacteria

1. Summary

Riboswitches are metabolite-binding domains within certain messenger RNAs that serve as precision sensors for their corresponding targets. Allosteric rearrangement of mRNA structure is mediated by ligand binding, and this results in modulation of gene expression. A class of riboswitches that selectively recognizes guanine and becomes saturated at concentrations as low as 5 nM are disclosed herein. In *Bacillus subtilis*, this mRNA motif is located on at least five separate transcriptional units that together encode 17 genes that are mostly involved in purine transport and purine nucleotide biosynthesis. These findings provide further examples of RNAs that sense metabolites and that control gene expression without the need for protein factors. Furthermore, it is now apparent that riboswitches contribute to the regulation of numerous fundamental metabolic pathways in certain bacteria.

2. Introduction

It is widely understood that the interplay of protein factors and nucleic acids guide the complex regulatory networks for genetic expression in modern cells. In most instances, protein factors appear to be well-suited agents for maintaining genetic expression networks. Proteins can adopt complex shapes and carry out a variety of functions that permit living systems to sense accurately their chemical and physical environments. Protein factors that respond to metabolites typically act by binding DNA to modulate transcription initiation (e.g., the lac repressor protein; Matthews, K. S., and Nichols, J. C., 1998, *Prog. Nucleic Acids Res. Mol. Biol.* 58, 127-164) or by binding RNA to control either transcription termination (e.g., the PyrR protein; Switzer, R. L., et al., 1999, *Prog. Nucleic Acids Res. Mol. Biol.* 62, 329-367) or translation (e.g., the TRAP protein; Babitzke, P., and Gollnick, P., 2001, *J. Bacteriol.* 183, 5795-5802). Protein factors respond to environmental stimuli by various mechanisms such as allosteric modulation or post-translational modifications, and are adept at exploiting these mechanisms to serve as highly responsive genetic switches (e.g., see Pavone, M., and Gann, A. (2002). *Genes and Signals* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.).

In addition to the widespread participation of protein factors in genetic control, it is now known that RNA can take an active role in genetic regulation. Recent studies have begun to reveal the substantial role that small non-coding RNAs play in selectivity targeting mRNAs for destruction, which results in down-regulation of gene expression (e.g., see Hannon, G. J. 2002, *Nature* 418, 244-251, and references therein). This process of RNA interference takes advantage of the ability of short RNAs to recognize the intended mRNA target selectively via Watson-Crick base complementation, after which the bound mRNAs are destroyed by the action of proteins. RNAs are ideal agents for molecular recognition in this system because it is far easier to generate new target-specific RNA factors through evolutionary processes than it would be to generate protein factors with novel but highly specific RNA binding sites.

Many studies have now confirmed that the complex three-dimensional shapes that some RNA molecules can mimic protein receptors and antibodies in their ability to selectively bind proteins or even small molecules (Gold, L., et al., 1995, *Annu. Rev. Biochem.* 64, 763-797; Hermann, T., and Patel, D., 2000, *Science* 287, 820-825). Furthermore, RNAs exhibit sufficient structural complexity to permit the formation of allosteric domains that undergo structural and functional modulation upon ligand binding (Soukup, G. A., and

[0445] The known riboswitches bind their target metabolites with high affinity and precision, which are essential characteristics for any type of molecular switch that can permit accurate and sensitive genetic control. For example, a recently identified riboswitch that responds to the coenzyme S-adenosylmethionine (SAM) binds it target with a dissociation constant \(K_{d}\) of ~4 nM (see Example 7). Furthermore, the riboswitch can discriminate ~100-fold against S-adenosylhomocysteine, which is a natural metabolite that differs from SAM by a single methyl group and an associated positive charge. Disclosed herein (Example 1) genetic control involving riboswitches is a widespread phenomenon with regard to its biological distribution and the target molecules that are being monitored. The observations that certain mRNAs from Archaeal organisms carry riboswitch-like domains (Storino, G. D., and et., Y., 2001, Proc. Natl. Acad. Sci. USA 98, 9465-9467; Rodionov, D. A., et al., 2002, J. Biol. Chem. 277, 48949-48959) and that several mRNAs from fungi and plants bind thiamine pyrophosphate (TPP) (Sudarsan, N., et al., 2003, RNA 9:644-647).

[0446] The genetic regulation of purine transport and purine biosynthesis pathways in bacteria, which are fundamental to the metabolic maintenance of nucleotides and nucleic acids (Switzer, R. L., et al., 2002, A. I. Sonenshein, et al., eds., ASM Press, Washington, pp. 255-269), were analyzed for the presence of riboswitches. In B. subtilis, numerous genes are involved in the biosynthesis of purines (pur operon with 12 genes; Ebbole, D. J., and Zalkin, H. 1987, J. Biol. Chem. 262, 8274-8287) and in the salvage of purine bases from degraded nucleic acids. The involvement of a regulatory protein factor has been proposed to participate in the control of the xpt-pbuX operon which encodes a xanthine phosphoribosyltransferase and a xanthine-specific purine permease, respectively (Christiansen, L. C., et al., 1997, J. Bacteriol. 179, 2540-2550). Although the protein factor PurK is known to serve as a repressor of transcription in the presence of elevated adenine concentrations (Weng, M., et al., 1995, Proc. Natl. Acad. Sci. USA 92, 7455-7459), no protein with corresponding function has been identified in B. subtilis that responds to guanine.

[0447] Disclosed herein the xpt-pbuX operon is controlled by a riboswitch that exhibits high affinity and high selectivity for guanine. This new found class of riboswitches is present in the 5‘-untranslated region (5‘-UTR) of five transcriptional units in B. subtilis, including that of the 12-gene pur operon. Thus, direct binding of guanine by mRNAs serves as a critical determinant of metabolic homeostasis for purine metabolism in certain bacteria. Furthermore, it was determined that the known classes of riboswitches, which respond to seven distinct target molecules, appear to control at least 68 genes in Bacillus subtilis that are of fundamental importance to central metabolic pathways. These findings indicate that riboswitches play a substantial role in metabolic regulation in living systems that direct interaction between small metabolites and RNA is a significant and widespread form of genetic regulation in bacteria.

[0448] 3. Experimental Procedures

[0449] i. Chemicals and Oligonucleotides

[0450] Guanine and its analogs xanthine, hypoxanthine, adenine, guanosine, 7-methylguanine, N2-methylguanine, 1-methylxanthine, 3-methylxanthine, 8-methylxanthine, 2-aminopurine, 2,6-diaminopurine, allopurinol, 2-amino-6-mercaptopurine, lumazine, and guanine-8-1H hydrochloride were purchased from Sigma. Inosine, uric acid, 2-amino-6-hromopurine, O-methyl guanine and pterin were purchased from Aldrich.

[0451] DNA oligonucleotides were synthesized by the Keck Foundation Biotechnology Resource Center at Yale University, purified by denaturing PAGE and eluted from the gel by crush-soaking in 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl, and 1 mM EDTA. Oligonucleotides were recovered from solution by precipitation with ethanol.

[0452] ii. Phylogenetic Analyses

[0453] G box domains were identified by sequence similarity to the xpt-pbuX 5‘-UTR by conducting a BLASTN search of Genbank using default parameters. These hits were expanded by searching for degenerate matches to the pattern (<<<[TA][6]<<<[2]ATNGG[2]<<<[5]GTTNCTAC[2]<<<[5]CCNNNAA[6]<<<[5]CC[8]<<<[5]) (SEQ ID NO:377). Angled brackets indicate base pairing. Bracketed numbers are variable gaps with constrained maximum length denoted. A total of four violations of this pattern were permitted when forming the phylogeny depicted in FIG. 23. It is important in this instance to note that only the BS3-xpt domain (that of the xpt-pbuX leader) has been shown to bind guanine. It was demonstrated that the molecular specificity of the VV1 representative is for adenine and not guanine (unpublished data). Given the possible trivial means by which a guanine-binding RNA aptamer might be altered to bind adenine (e.g. a C to U change if the C residue is used by the aptamer to make a Watson-Crick-pairing interaction with guanine), it cannot be ruled out that other representatives also have altered molecular recognition.

[0454] iii. In-Line Probing of RNA Constructs

[0455] The B. subtilis 201 xpt leader and truncated 93 xpt aptamer RNAs were prepared by in vitro transcription using T7 RNA polymerase and the appropriate PCR DNA templates, and were subsequently 5‘-[32P]-labeled using a protocol similar to that described previously (Seetharaman, S. et al., 2001, Nature Biotechnol. 19, 336-341). Labeled precursor RNAs (~2 nM) were subjected to in-line probing using conditions similar to those described in Example 2. Reactions (10 µl) were incubated for 40 hr at 25°C in a buffer containing 50 mM Tris (pH 8.5 at 25°C), 20 mM MgCl2, and 100 mM KCl in the presence or absence of purines as indicated for each experiment. Purine concentrations ranging from 1 nM to 10 µM were typically employed but ranged as high as 300 µM for poor-binding ligands. Denaturing 10% PAGE was used to separate spontaneous cleavage products and a Molecular Dynamics PhosphorImager was used to view the results. Quantitation of spontaneous cleavage yields was achieved by using ImageQuant software. Since concentrations of RNA below 2 nM for in-line probing cannot be used easily due to insufficient levels of signal, apparent \(K_{D}\) values near this concentration reflect the maximum possible value.
[0456] iv. Equilibrium Dialysis

[0457] Equilibrium dialysis assays were conducted using a DispoEquilibrium Dialyzer (ED-1, Harvard Bioscience), wherein two chambers a and b were separated by a 5,000 MWCO membrane. The final composition of buffer included 50 mM Tris-Cl (pH 8.5 at 25°C), 20 mM MgCl₂ and 100 mM KCl (30 μL delivered to each chamber). Chamber a also contained 100 nM ³H-guanine, while chamber b also contained 300 nM of xpt RNA constructs as indicated for each experiment. After 10 hr of equilibrium at 25°C, a 5 μl aliquot from each chamber was removed for quantitation by liquid scintillation counter. When appropriate, an additional 5 μL of buffer was added to a and an equivalent volume of buffer containing 500 nM unlabeled purine was added to b. After an additional 10 hr incubation at 25°C, 5 μl aliquots were again drawn for quantitation of tritium distribution.

[0458] v. Construction of xpt-lacZ Fusions

[0459] Genetic manipulations were conducted using approaches similar to those described elsewhere herein. Briefly, a DNA construct encompassing nt -121 to +197 relative to the transcription start site of the xpt-pbuX operon from B. subtilis strain 1A40 (Bacillus Genetic Stock Center, Columbus, Ohio) was PCR amplified as an EcoRI-BamHI fragment. The product was cloned into pDG1661 at a site directly upstream of the lacZ reporter gene. Mutants were created within the engineered pDG1661 by using the appropriate primers and the QuickChange Site-directed mutagenesis kit (Stratagene). Plasmid variants were integrated into the amyE locus of strain 1A40. Transformants were selected for chloramphenicol (5 μg/mL) resistance and screened for sensitivity to spectinomycin (100 μg/mL). The integrity of each construct was confirmed by sequencing.


[0461] B. subtilis cells were grown with shaking at 37°C in minimal medium containing 0.4% w/v glucose, 20 g L⁻¹ (NH₄)₂SO₄, 25 g L⁻¹ K₂HPO₄, 3H₂O, 6 g L⁻¹ KH₂PO₄, 1 g L⁻¹ sodium citrate, 0.2 g L⁻¹ MgSO₄·7H₂O, 0.2% glutamate, 5 μg mL⁻¹ chloramphenicol, 50 μg mL⁻¹ L-tryptophan, 50 μg mL⁻¹ L-lysine and 50 μg mL⁻¹ L-methionine. Purines were added at a final concentration of 0.5 μg mL⁻¹. Cells at mid exponential stage (A₅₅₀ of ~0.1) were harvested by centrifugation and resuspended in minimal media in the absence or presence of a purine from 0.5 μg mL⁻¹ to 200 μg mL⁻¹ for 30 min at 37°C. Although the poor solubility of guanine causes the formation of a detectable level of precipitate at this concentration, no adverse effects of cell growth were observed. Unless otherwise specified, cells were incubated for an additional 3 hrs before performing β-galactosidase assays. Data presented in FIG. 28C was generated as described above with the exception that β-galactosidase assays were performed at the times indicated.

[0462] 4. Results and Discussion

[0463] i. A Conserved Domain in the 5'-UTR of Several B. subtilis mRNAs.

[0464] The xpt-pbuX operon is regulated by guanine, hypoxanthine, and xanthine. These purine compounds share chemical similarity and are adjacent to each other in the pathways of purine salvage. In contrast to the pur operon, regulation of the xpt-pbuX operon remains unaffected by adenine in a strain wherein adenine deaminase is inactive (Christiansen, L. C., et al., 1997, J. Bacteriol. 179 2540-2550). These observations had fostered speculation that an unidentified protein factor might be involved in guanine recognition (Ebbelle, D. J., and Zalkin, H. 1987, J. Biol. Chem. 262, 8274-8287), however, such a genetic factor has not been identified. Moreover, the 5'-UTR of the xpt-pbuX mRNA is rather large (185 nucleotides), which could be sufficient to accommodate a riboswitch domain.

[0465] Riboswitches are typically composed of two functional domains: an aptamer that selectively binds its target metabolite and an expression platform that responds to metabolite binding and controls gene expression by allosteric means. The most conserved portion of known riboswitches is the aptamer domain, whereas the adjoining expression platform can vary widely in sequence and in secondary structure. The high sequence conservation of the aptamer is due to the fact that the RNA must retain its ability to form a receptor for a chemical that does not change through evolution. In contrast, the expression platform can form one of a great diversity of structures that permit genetic control in response to ligand binding by the aptamer domain. This evolutionary conservation was exploited to design a database search for xpt-pbuX 5'-UTR sequences that are present in other B. subtilis genes and also in other bacterial species. Five transcriptional units within B. subtilis that closely correspond in sequence and predicted secondary structure with nucleotides 14 through 82 of the xpt-pbuX 5'-UTR (FIG. 23) were identified. A total of 32 representatives of this domain were identified amongst several Gram-positive and Gram-negative bacteria. Other members can exist as well.

[0466] From this representative set of RNAs, a consensus sequence and secondary structure for the conserved RNA motif termed the “G box” (FIG. 24A) were identified. The secondary structure of the G box is composed of a three-stem (P1 through P3) junction, wherein significant sequence conservation occurs within P1 and in the unpaired regions. Furthermore, it was found that stems P2 and P3 both favor seven base pairs in length with one- or two-base mismatches permitted. This unusual conservation of stem length implies that these structural elements establish distance and orientation constraints of their stem-loop sequences relative to the three-stem junction. Some base-pairing potential exists between the two stem-loop sequences, which might permit the formation of a pseudoknot. These characteristics indicate that G-box domains most likely use conserved secondary- and tertiary-structure elements to adopt a precise three-dimensional fold.

[0467] ii. The G Box RNA from the xpt-pbuX 5'-UTR of B. subtilis Binds Guanine

[0468] Two RNA constructs based on the xpt-pbuX 5'-UTR of B. subtilis were prepared to examine whether the mRNA selectively binds guanine or its closest analogs. A double-stranded DNA template corresponding to the entire 5' UTR and the first four codons of the xpt-pbuX mRNA was generated by PCR using primers that introduced a promoter sequence for T7 RNA polymerase and several nucleotide additions and mutations that permit further manipulation (FIG. 24B; see also Experimental Procedures). A truncated form of this construct also was created by PCR that encompasses the 5' half of the UTR. Upon transcription, the shorter DNA template generates a 93-nucleotide transcript termed 93 xpt, while the longer template produces a 201-nucleotide transcript termed 201 xpt.

[0469] These precursor RNAs were 5'32P-labeled and subjected to an in-line probing assay (e.g., see Example 1) wherein the spontaneous cleavage of RNA linkages within an aptamer is monitored in the presence and absence of its corresponding ligand. It was found that the patterns of sponta-
neous cleavage of the 93 xpt (FIG. 24C) and the 201 xpt (FIG. 25A) RNAs undergo significant alteration upon addition of guanine at a concentration of 1 μM. Both hypoxanthine and xanthine also induce modulations of spontaneous cleavage at this concentration. Specifically, four major regions exhibit ligand-mediated reduction in spontaneous cleavage (FIGS. 24B and 24C). However, the presence of 1 μM adenine (and as much as 1 mM) does not alter the pattern of RNA cleavage products. These results indicate that the G box domain in the 5' UTR of the B. subtilis xpt-pbxX mRNA serves as an aptamer for guanine and related purines, and that this aptamer undergoes significant structural modulation upon ligand binding. In the context of a riboswitch, this allosteric function could be harnessed by the mRNA to modulate structural elements that regulate gene expression.

[0470] In a preliminary assessment of the affinity that the guanine aptamer has for its target, in-line probing with 201 xpt in the presence of various concentrations of guanine was conducted. As expected, increasing concentrations provided progressively decreasing amounts of spontaneous cleavage at the four major sites of structural modulation (FIG. 25A). Half-maximum levels of modulation were observed when a concentration of ~5 nM guanine is used for in-line probing (FIG. 25B). Although this implies that the Kd for 201 xpt under these conditions is ~5 nM, it is important to note that the actual value might be somewhat lower because of the limitations of the in-line probing assay (see Experimental Procedures). In addition, the Kd was determined under non-physiological conditions (e.g. high Mg2+ and elevated pH), and so the binding affinity might be somewhat different in vivo. However, using this number for comparison, the affinity of the 201 xpt RNA for guanine is more than 10,000-fold greater than that of the Tetrahymena group I ribozyme for its guanosine monophosphate substrate (McConnell, T. S., et al., 1993, Proc. Natl. Acad. Sci. USA 90, 8362-8366). This difference most likely reflects the relative differences in concentrations of the two compounds that the RNAs experience inside their respective cellular environments.

[0471] iii. The Guanine Aptamer Discriminates Against Many Purine Analogues

[0472] To maintain precise metabolic homeostasis, the cell must be able to sense the concentration of its target metabolite, but also must prevent regulatory cross talk with other compounds that otherwise might inadvertently trigger genetic modulation. Indeed, a hallmark of other riboswitches is the ability to discriminate between closely related metabolites. For example, the FMN and TPP riboswitches discriminate against the unphosphorylated coenzyme precursors thiamine and riboflavin by ~1,000 fold (see Examples 2 and 3).

[0473] This requirement for obligate molecular discrimination against related metabolites is expected to be extreme with guanine riboswitches, as there are numerous purine nucleosides and nucleotides, purine bases, and purine-like compounds that are present in the cell. Using the in-line probing strategy described in FIG. 25, the apparent Kd values of the 93 xpt RNA were established for a variety of purines and purine analogs. Hypoxanthine and xanthine exhibit Kd values that are closest to the value determined for guanine, while adenine has a Kd value in excess of 300 μM (FIG. 26A). These results are consistent with the observation that adenine does not significantly repress expression of the xpt-pbxX operon as do the other purines (Christiansen, L. C., et al., 1997, J. Bacteriol. 179, 2540-2550). However, it is not clear whether hypoxanthine and xanthine might repress gene expression by directly binding a guanine riboswitch, or whether they might first be converted into guanine before influencing genetic control.

[0474] It was found that alteration of every functionalized position on the guanine heterocycle causes a substantial loss of binding affinity (FIG. 26B, FIG. 27). For example, the oxygen atom at position 6 of guanine is a significant determinant of molecular recognition, as demonstrated by the losses in apparent Kd for 2-aminopurine (~10,000-fold loss), 2-amino-6-bromopurine (~1,000 fold), and O-methylguanine (~100 fold). Most molecular interactions could be explained by invoking hydrogen bonding contacts between the RNA and guanine with the exception of the molecular interaction at C8. Here, presumably the RNA structure creates a steric clash with analogs that carry additional bulk, such as 8-methylxanthine (~10,000 fold) and uric acid (~10,000 fold).

[0475] A summary of the likely molecular recognition features that the guanine aptamer requires for maximum affinity is depicted in FIG. 26C. However, the likely possibility that significant binding affinity could be derived through base stacking was not examined. The presence of so many productive contacts between the RNA and all faces of guanine suggest that the ligand is most likely entirely engulfed by the aptamer's structure. This would also explain why the RNA is capable of generating recognition via steric occlusion of bulkier compounds such as uric acid. In certain biological environments, for example, uric acid can build up to high concentrations that permit crystallization. In such environments, a bacterium would require a high level of discrimination to prevent undesirable repression of guanine-regulated genes. In light of such molecular recognition challenges, it is not surprising that an RNA genetic switch would evolve extensive molecular contacts with its target compound.

[0476] iv. Confirmation of Guanine Aptamer Function by Equilibrium Dialysis

[0477] Equilibrium dialysis was used to provide further evidence that the G box RNA from the xpt-pbxX operon binds guanine preferentially over other purines and purine analogs. A substantial shift in titrated guanine is expected to occur in a two-chamber dialysis apparatus when an excess of functional RNA is added to one chamber (FIG. 27A). Furthermore, this shifted equilibrium should return to unity upon addition of an excess of unlabeled competitor ligand. As expected, it was observed that greater than 90% of titrated guanine co-localizes with 93 xpt RNA, and subsequently redistributes when an excess of unlabeled guanine is introduced. In contrast, the presence of excess unlabeled analogs has no effect on co-localization of 3H-guanine and the RNA (FIG. 27B). Even the nucleoside guanosine (9-ribosylguanine) fails to restore equal distribution of guanine between the two chambers, which is consistent with the RNA folding to form a tight pocket for the base alone.

[0478] Both in-line probing and equilibrium dialysis data indicate that this natural aptamer binds guanine with high affinity and specificity. In a previous study, in vitro evolution was used to isolate a purine-binding aptamer from a pool of random-sequence RNAs (Kiga, D., et al., 1998, Nucleic Acids Res. 26, 1755-1760). This engineered aptamer exhibits a Kd of 1.3 μM for guanine, and shows only a 2- to 3-fold discrimination against hypoxanthine and xanthine. The lower specificity and affinity of this aptamer for selected purines is due to the fact that only the N1, N7 and O6 positions are important for molecular recognition. In contrast, the G box
RNA appears to make productive contacts with all available functional groups on guanine, presumably through hydrogen bonding (FIG. 26C).

[0479] A variety of mutations were introduced into the G box domain to examine the importance of several structural elements and conserved nucleotides (FIG. 28A). The influence of these mutations on guanine binding was determined in the context of the 93 xpt RNA by using equilibrium dialysis. Mutations that independently disrupt the three stems (M1, M4 and M6) cause a loss of binding function, as does a variant RNA (M3) that carries two mutations in the central junction (FIG. 28B). In contrast, the effects of the disruptive stem mutations are largely reversed by making compensatory mutations (M2, M5 and M7) that restore base pairing. These results are consistent with the phylogenetic analysis (FIG. 23), which indicates that stem structure is important but that the precise sequence composition of these elements is of less importance.

[0481] Binding function of variant aptamers in vitro also correlates with genetic control in vivo. The results disclosed herein confirmed earlier findings that a reporter gene carrying the 5'-UTR of the xpt-pbuX mRNA is repressed by guanine, and to a lesser extent by hypoxanthine and xanthine (Christenssen, T. C., et al., 1997, J. Bacteriol. 179, 2540-2550). Specifically, transcriptional fusions were created between a β-galactosidase reporter gene and variant xpt-pbuX 5'-UTR sequences carrying the mutations described in FIG. 28A. B. subtilis chromosomal transfectants using the wild-type sequence exhibit the expected levels of genetic modulation (FIG. 28C). Although the xpt aptamer exhibits dissociation constants for xanthine and hypoxanthine that are essentially identical in vitro, the differences in genetic modulation by these compounds in vivo might be due to differences in their cellular concentrations.

[0482] Aptamer variants with impaired guanine binding in vitro also exhibit a loss of β-galactosidase repression (FIG. 28D). Furthermore, restoration of base pairing in stems P1 through P3 results in restored genetic control. The M2 variant is of particular interest because it not only exhibits restored genetic control, but also provides modest expression of β-galactosidase in the absence of guanine Riboswitch function requires the action of an aptamer for molecular sensing as well as an expression platform that transduces RNA-ligand complex formation into a genetic response. Examples of TPP and FMN riboswitches (see Examples 2 and 3) appear to function by differential formation of terminator and antiterminal structures. Such ligand-induced formation of transcription anti-termination structures also appears to be the basis of expression platform mechanisms used by numerous SAM riboswitches (see Example 7). Construct M2 carries three mutations within the putative ant-terminator structure of the xpt-pbuX leader, and thus is expected to exhibit an overall reduction of reporter gene expression because these mutations should bias structure folding towards terminator stem formation.

[0483] The results of these mutational and functional analyses confirm the major features of the secondary structure model (P1 though P3) and demonstrate that they are critical for metabolite binding. Furthermore, the correlation between ligand binding and genetic control indicates that the G box and adjacent nucleotides of the xpt-pbuX leader sequence operate in concert to function as a guanine-dependent riboswitch, most likely by operating via allosteric control of transcription termination.

[0484] vi. Riboswitches Control Fundamental Biochemical Pathways

[0485] Our findings indicate that the G box RNA of the xpt-pbuX operon is a key structural element of a guanine-sensing riboswitch that exhibits extraordinary affinity and selectivity for its target. In B. subtilis, this general riboswitch motif appears to control at least five transcriptional units (FIG. 23). Although the precise function of several of the gene products in this newly identified regulon have not been clearly defined, the known genes from B. subtilis and from other organisms are mostly related to purine metabolism. Based on the results disclosed herein, it is likely the G box domain within the 5'-UTR of this large pur operon is responsible for guanine-dependent riboswitch regulation, and that the genetic regulatory mechanism might be similar to that proposed herein for the xpt-pbuX operon.

[0486] The distribution of G box domains in B. subtilis and other bacteria suggests that this class of metabolite-binding RNAs controls a regulon that is essential for cell survival. In B. subtilis, guanine riboswitches (or related adenine-dependent riboswitches—see the legend to FIG. 23) appear to provide at least some contribution to the genetic regulation of 17 genes. The discovery of guanine-dependent riboswitches adds to a growing list of similar metabolite-sensing RNAs. For example, a class of riboswitches that responds to SAM (McDaniel, B. A., et al., 2003, Proc. Natl. Acad. Sci. USA 100, 3083-3088; Epshtein, V., et al., 2003, Proc. Natl. Acad. Sci. USA 100, 5052-5056) controls a regulon of as many as 26 genes that are involved in coenzyme biosynthesis, amino acid metabolism, and sulfur metabolism. When included with genes that are controlled by other riboswitch classes, at least 68 genes (nearly 2% of its total genetic complement) are under riboswitch control (FIG. 29).

[0487] Riboswitches for ligands such as guanine and SAM apparently are serving as master control molecules whose concentrations are being monitored to ensure homeostasis of a much wider set of metabolic pathways. Riboswitches also seem to permit metabolite surveillance and genetic control with the same level of precision and efficiency as that exhibited by protein factors. Therefore, these RNA switches could have emerged late in the evolution of modern biochemical architectures because they are functionally comparable to genetic switches made of protein. However, given their fundamental role in metabolic maintenance and the widespread phylogenetic distribution of certain riboswitches, it is consistent that aptamer domains similar to these might have been the primary mechanism by which RNA-world organisms detected metabolites and controlled biochemical pathways before the emergence of proteins.

[0488] 5. Conclusions

[0489] This demonstration that guanine is sensed by metabolite-binding mRNAs expands the known classes of riboswitches, and provides additional evidence that certain bacterial RNAs are responsible for monitoring the concentrations of critical coenzymes and other compounds that are fundamental to all living systems. Phylogenetic analyses and biochemical data indicate that many bacteria and, in some instances, eukaryotes (Sudarsan, N., et al., 2003, RNA 9:644-647) entrust riboswitches to sense essential metabolites and mediate genetic control. Although protein factors undoubtedly could be used to carry out these important regulatory tasks, based on the disclosure herein, highly structured RNAs...
are well suited for this role. If RNA polymers were a poorly suited medium for generating metabolite receptors with high affinity and precision, then one would expect that evolution would have long ago replaced them by protein factors.

**[0490]** Disclosed herein it is consistent (e.g. see Examples 1 and 2) that riboswitches are derivatives of an ancient genetic control system that monitored metabolic and environmental signals before the evolutionary emergence of proteins. Interestingly, each of the metabolite targets of riboswitches has been proposed to come from an RNA world (White, H. B. III, 1976, J. Mol. Evol. 7, 101-104; Benner, S. A., et al., 1989, Proc. Natl. Acad. Sci. USA 86, 7054-7058; Jeffares, D. C., et al., 1998, J. Mol. Evol. 46, 18-36; Jadhav, V. R., and Yanus, M., 2002, Biochimie 84, 877-888). The identification of guanine as a trigger for riboswitches is consistent with metabolite sensing RNAs having originated very early in evolution. Also disclosed herein is another class of riboswitches that responds to the amino acid lysine (FIG. 29). Although all riboswitches could be more recent evolutionary inventions, even the origin of the lysine riboswitch might date from before the last common ancestor and back to a time when living systems were transitioning from a pure RNA world to a more modern metabolic state that made use of encoded protein synthesis.

**G. Example 7**

**S-Adenosylmethionine Riboswitch**

**[0491]** Riboswitches are metabolite-binding RNA structures that serve as genetic control elements for certain messenger RNAs. These RNA switches have been identified in all three kingdoms of life and are typically responsible for the control of genes whose protein products are involved in the biosynthesis, transport, or utilization of the target metabolite. Disclosed herein, is a highly conserved RNA domain found in bacteria serves as a riboswitch that responds to the coenzyme S-adenosylmethionine (SAM) with remarkably high affinity and specificity. SAM riboswitches undergo structural reorganization upon introduction of SAM, and these allosteric changes regulate the expression of 26 genes in *Bacillus subtilis*. This and related findings indicate that direct interaction between small metabolites and allosteric mRNAs is a significant and widespread form of genetic regulation in bacteria.

**[0492]** 1. Results

**[0493]** i. Identification of a SAM-Responsive Riboswitch

**[0494]** Each of the compounds sensed by previously identified riboswitches (coenzyme B_{12}, TPP, FMN) is used as a coenzyme by modern protein enzymes. Interestingly, these coenzymes have significant structural similarity to RNA, which has been used to support speculation that they might also have been used as coenzymes by ancient ribosomes in an RNA world (S. A. Benner, et al., Proc. Natl. Acad. Sci. USA 86, 7054 (1989); H. B. White III, J. Mol. Evol. 7, 101 (1976); D. C. Jeffares, et al., J. Mol. Evol. 46, 18 (1998). If modern riboswitches are direct descendents of RNA control systems that originated in the RNA world, then the metabolites they sense and the metabolic pathways that they control will be of fundamental importance to modern biochemical processes. To further assess this hypothesis, a search for additional riboswitches, to determine their biochemical characteristics, and to establish their role in genetic control on a genome-wide level was performed.

**[0495]** In this effort the S box was examined (E. J. Grundy, T. M. Henkin, Mol. Microbiol. 30, 737 (1998)), which is a highly conserved sequence domain (FIG. 30A) that is located within the 5'-untranslated region (5'-UTR) of certain messenger RNAs in Gram-positive bacteria. Both genetic and sequence analyses suggest that the S box domain serves as a genetic control element for a regulon composed of 11 transcriptional units. These mRNAs encode as many as 26 different genes in *B. subtilis* that are involved in sulfur metabolism, methionine biosynthesis, cysteine biosynthesis, and SAM biosynthesis. However, the nature of the putative regulatory factor and the metabolite to which it respond has not been established (T. M. Henkin, Curr. Opin. Microbiol. 3, 49 (2000); F. J. Grundy, T. M. Henkin, Frontiers Biosci. 8, D20 (2003)). An RNA construct corresponding to the first 251 nucleotides of the yijJ mRNA of *B. subtilis* (FIG. 30B) was prepared by in vitro transcription (G. A. Soukup, R. R. Breakey, RNA 5, 1308 (1999)). The yijJ gene product is a putative methylene tetrahydrofolate reductase—an enzyme proposed to be involved in methionine biosynthesis (F. J. Grundy, T. M. Henkin, Mol. Microbiol. 30, 737 (1998)). The 251 yijJ RNA was subjected to "in-line probing", which reveals locations of structured and unstructured portions of RNA polymers by relying on the variability in rates of spontaneous RNA phosphodiester cleavage caused by differences in structural context. In-line probing may also reveal nucleotides participating in metabolite-induced structural modulation (see Examples 1-3).

**[0496]** Whether the 251 yijJ RNA might bind S-adenosylmethionine (SAM) was analyzed. Indeed, upon separation by polyacrylamide gel electrophoresis (PAGE), the pattern of spontaneous RNA cleavage products (FIG. 30C) was indicative of a highly structured RNA element that undergoes conformational modulation upon introduction of SAM to a final concentration of either 0.1 mM or 1 mM. In contrast, no structural modulation was evident upon the introduction of methionine at the same concentrations, suggesting that the RNA might require both the methionine and 5'-deoxyadenosyl moieties of SAM to induce structural reorganization. The locations of the ligand-induced modulations (FIG. 30D) indicated that the conserved core of the S box RNA serves as a natural aptamer (L. Gold, et al., Annu. Rev. Biochem. 64, 765 (1995)), for SAM. Similar results were observed with 124 yijJ, which encompasses nucleotides 28 through 149 of the mRNA leader plus two G residues at the 5' terminus.

**[0497]** ii. Molecular Recognition by a SAM-Dependent Riboswitch

**[0498]** A genetic switch that responds to metabolites must be able to bind its target with a dissociation constant (K_d) that is relevant to physiological concentrations. Furthermore, the metabolite receptor must be able to discriminate precisely against closely related compounds that are likely to occur in the same milieu, or risk undesirable modulation of gene expression. Therefore, the affinity of the yijJ RNA for SAM was assessed, and the ability of the RNA to discriminate against biologically relevant compounds that are structurally similar to this target (FIG. 31A).

**[0499]** The K_d of 251 yijJ for SAM was determined by using in-line probing to monitor the extent of structural modulation over a range of ligand concentrations (FIG. 31B, left). Although the K_d of 251 yijJ for SAM is ~200 nM, the minimized aptamer domain represented by 124 yijJ exhibits a K_d of ~4 nM under the dissected assay conditions. Such improvements in binding affinity by minimized aptamer domains have been observed (see Example 2). This most likely reflects greater structural preorganization of the ligand binding form of the aptamer domain due to the elimination of
the adjoining expression platform, which otherwise would permit alternative folding to occur. Tight binding was also observed when the 124 yiT RNA was interrogated by using a Scatchard analysis with tritiated SAM. The assessment of binding affinity indicated that the $K_a$ for the 124 yiT aptamer is more than 1000-fold improved compared to that reported recently for a related RNA (McDaniel, B. et al., *Proc. Natl. Acad. Sci. USA* 100, 3083-3088 (2003)). Normal concentrations of SAM in bacteria are typically in the low micromolar range (McDaniel, B. et al., *Proc. Natl. Acad. Sci. USA* 100, 3083-3088 (2003)), however, most of this coenzyme pool is probably bound by enzymes. Therefore the low $K_a$ exhibited by this riboswitch might be needed to sense the concentration of free SAM.

As expected, the 124 yiT RNA achieves a high level of molecular discrimination against analogs of SAM. For example, the RNA exhibits ~100-fold discrimination against SAH (Fig. 31b, right), which is produced upon utilization of SAM as a coenzyme for methylation reactions (F. Takusa-gawa, et al., In: *Comprehensive Biological Catalysis*, M. Sinnott, ed., Academic Press, Vol. 1, pp. 1-30 (1998)). Thus, the aptamer must form a binding pocket for SAM that can sense the absence of a single methyl group and an associated loss of positive charge. Similarly, the RNA discriminates nearly 10,000-fold against SAC, which is another biological compound that differs from SAH by the absence of a single methylene group. This pattern of molecular discrimination was confirmed by using equilibrium dialysis (Fig. 31c).

As expected, the 124 yiT aptamer RNA binds each of the three SAMs with high affinity, with no significant sequence specificity. The binding of each SAM to the aptamer is reversible and can be competed by the presence of other SAM analogs. The 124 yiT RNA binds SAM with an apparent dissociation constant ($K_a$) of ~1 nM, which is comparable to the dissociation constant for the 124 yiT aptamer bound to the wild-type 5S rRNA (McDaniel, B. et al., *Proc. Natl. Acad. Sci. USA* 100, 3083-3088 (2003)).

The mechanism of SAM-induced termination (Fig. 33b) most likely involves the ligand-mediated formation of alternative hairpin structures that permit transcriptional read-through (anti-terminator formation without SAM) or that cause termination (terminator formation with SAM). This mechanism was examined by generating several mutant metL constructs that carry disruptive or compensatory changes in the expression platform (Fig. 33b). SAM causes an additional ~20% yield in transcription termination in a mutant (Mab) that carries six mutations relative to the wild-type metL riboswitch, which retains proper terminator and anti-terminator base complementation. However, incomplete representation of these six mutations that do not permit normal pairing interactions to occur permits little or no SAM-mediated transcription modulation. Furthermore, mutations that disrupt terminator stem formation (Ms) yield lower levels of termination, while mutations that disrupt anti-terminator stem formation (Mab, Me) yield higher levels of termination (Fig. 33b). These findings indicate that the RNA structural modulation induced by SAM binding mediates genetic control by sequestering an anti-terminator sequence, and thus favors the formation of a transcriptional terminator hairpin.
from the same organism (FIG. 34a). However, the cysH homolog from B. anthracis exhibits a Kₜ that matches that of yiiA (FIG. 34b), implying that the B. subtilis cysH aptamer has suffered one or more mutations that have somewhat degraded binding affinity.

**[0510]** 2. Conclusion

**[0511]** Current biochemical and bioinformatics data indicate that B. subtilis has at least 68 genes (nearly 2% of its total genetic complement) under riboswitch control. Moreover, each of these mRNAs is responding to biological compounds that are universal in biology. The fact that genetic control elements for fundamental metabolic processes are formed by RNA indicates that this polymer has the structural sophistication needed to precisely monitor chemical environments and transduce metabolic binding events into genetic responses. A more detailed analysis of riboswitch structures at the atomic level would be of great utility in determining how metabolic binding promotes allosteric reorganization RNA genetic switches.

**[0512]** Riboswitches for ligands such as SAM and guanine appear to be serving as master control molecules whose concentrations are being monitored to ensure homeostasis of a much wider set of metabolic pathways. Riboswitches seem to permit metabolic surveillance and genetic control with the same level of precision and efficiency as that exhibited by protein factors, and thus could have emerged late in the evolution of modern biochemical architectures.

**[0513]** 3. Methods

**[0514]** i. DNA Oligonucleotides and Chemicals

**[0515]** Synthetic DNAs were purchased from The Keck Foundation Biotechnology Resource Center at Yale University. Preparation of RNAs by in vitro transcription was conducted (Seetharaman, S., et al., Nat. Biotechnol. 19, 336-341 (2001)) and the products were purified as described in Example 2. SAM, various analogs of SAM, and S-adenosyl-L-methionine-methyl-³H (³H-SAM) were purchased from Sigma.

**[0516]** ii. DNA Constructs

**[0517]** A yiiA DNA construct encompassing nucleotides -380 to +15 relative to the translation start site was constructed using primers that generated EcoRI and BamHI restriction sites upon PCR amplification of B. subtilis chromosomal DNA (strain 168). The product was cloned into pDG1661 (ref 26; Bacillus Genetic Stock Center, Columbus, Ohio) using these restriction sites, which places the riboswitch immediately upstream of the lacZ reporter gene. Mutants were created by using the appropriate mutagenic primers and the QuickChange site-directed mutagenesis kit (Stratagene). All sequences were confirmed by sequencing.

**[0518]** iii. In Vivo Analysis of Riboswitch Function

**[0519]** B. subtilis strain 1A234 was obtained from the Bacillus Genetic Stock Center, Columbus, Ohio. Cells were grown with shaking at 37° C. either in rich media (2XYT broth or tryptose blood agar base) or defined media (0.5% w/v glucose, 20 g L⁻¹ (NH₄)₂SO₄, 183 g L⁻¹ KH₂PO₄, 3H₂O, 60 g L⁻¹ K₂HPO₄, 10 g L⁻¹ sodium citrate, 2 g L⁻¹ MgSO₄, 7H₂O, 5 µM MgCl₂, 50 µg L⁻¹ triphenylphosphite, and 50 µg L⁻¹ glutamate, Methionine was added to 50 µg L⁻¹ for routine growth. Growth under methionine-limiting conditions was established by incubation under routine growth conditions to an A₅₅₀ of 0.1, at which time the cells were pelleted by centrifugation, resuspended in minimal media, split into two aliquots, and supplemented with either 50 µg L⁻¹ (+methionine) or 0.75 µg L⁻¹ (−methionine) (FIG. 32). Cultures were incubated for an additional 3 hr before performing β-galactosidase assays. Transformations of pDG1661 variants (see DNA constructs) into B. subtilis were performed as described elsewhere (H. Jarmer, et al., FEMS Microbiol. Lett. 206, 197 (2002)). The correct transformants were identified by selecting for chloramphenicol (5 µg mL⁻¹) resistance and screening for spectinomycin (100 µg mL⁻¹) sensitivity. Proper site-specific genomic insertion by double cross-over recombination was confirmed by PCR using amyE-specific primers.

**[0520]** iv. In Vitro Transcription Termination Assays

**[0521]** Transcription reactions (10 µL) containing ~30 pmoles of specific template DNA, 200 µM each NTP, 5 µCi [α-³²P]UTP (1 Ci = 37 GBq) and 50 units of T7 RNA polymerase (New England Biolabs) were incubated in the presence of 50 mM Tris-HCl (pH 7.5 at 23° C.), 15 mM MgCl₂, 2 mM spermidine, 5 mM DTT at 37° C. for 2 hr. SAM and its analogs were added to a final concentration of 50 µM. Transcription templates were generated for all 11 riboswitch domains in the S box region of B. subtilis by using PCR with corresponding primers that in each case produced transcripts beginning with GG, encompassing the putative natural transcription start (J. F. Grundy, T. M. Henkin, Mol. Microbiol. 30, 737 (1998)), and including the first 13 codons of the adjoining open reading frame. Transcription products were separated by denaturing 6% PAGE and visualized by Phosphorimager. Termination yields were approximated by determining the ratio of RNAs in the termination band relative to the combined terminated and full-length RNAs.

H. Example 8

**Adenine Riboswitches**

**[0522]** A class of riboswitches that recognizes guanine and discriminates against most other purine analogs was recently identified (see Example 6). Representative RNAs that carry the consensus sequence and structural features of guanine riboswitches are located in the 5'-untranslated region (UTR) of numerous genes of prokaryotes, where they control expression of proteins involved in purine salvage and biosynthesis. This example shows that three representatives of this phylogenetic collection bind adenine with values for apparent dissociation constant (approx. Kₜ) that are several orders of magnitude better than for guanine. The preference for adenine is due to a single nucleotide substitution in the core of the riboswitch, wherein each representative most likely recognizes its corresponding ligand by forming a Watson/Crick base pair. In addition, the adenine-specific riboswitch associated with the ydhL gene of Bacillus subtilis functions as a genetic ‘ON’ switch, wherein adenine binding causes a structural rearrangement that precludes formation of an intrinsic transcription terminator stem.

**[0523]** Guanine-sensing riboswitches are a class of RNA genetic control elements that modulate gene expression in response to changing concentrations of this compound (see Example 6). This is one of a number of classes of metabolite-binding riboswitches that regulate gene expression in response to various fundamental compounds such as lysine and the coenzymes fMN, SAM, B₃₄ and TPP (thiamin pyrophosphate) (see Example 6). Typically, each riboswitch is composed of two functional domains, an aptamer and an expression platform, that function together as a transducer of chemical signals into altered patterns of gene expression. The aptamer serves as a specific receptor for the target metabolite,
wherein ligand binding brings about allosteric changes in both the aptamer and expression platform domains.

Detailed examinations of the ligand specificities for the natural aptamers from guanine- and lysine-specific riboswitches have been conducted (see Example 6), and less comprehensive examinations of the FMN, SAM, B12 and TPP aptamers have been conducted (see Examples 1-3). In each case, the RNAs exhibit high levels of molecular discrimination by disfavoring the binding of even closely related metabolite analogs. This characteristic of high molecular discrimination is a hallmark of enzymes and receptors, including genetic regulatory factors, which need to carry out biological processes with great precision in the presence of complex chemical mixtures.

The molecular recognition characteristics of guanine riboswitches are distinguished by the fact that nearly every position around the purine heterocycle appears to be critical for high affinity binding by the aptamer. Thus, the arrangement of the binding pocket permits the riboswitch to control gene expression in response to changing guanine concentrations, but prevents modulation of gene expression in response to increasing concentrations of adenine (see Example 6; Cristiansen, L. C., et al., J. Bacteriol. 179, 2540-1550 (1997)). However, it is likely that receptors made of RNA, like their protein counterparts, could acquire altered molecular recognition characteristics as a result of natural selection. This would permit riboswitches to emerge through evolution that selectively sense and respond to metabolites that are proximal in metabolic pathways.

This example confirms the existence of a variant class of riboswitches that responds to adenine. These riboswitches carry an aptamer domain that corresponds closely in sequence and secondary structure to the guanine aptamer described recently (see Example 6). However, each representative of the adenine sub-class of riboswitches carries a C to U mutation in the conserved core of the aptamer, indicating that this residue is involved in metabolite recognition. The results indicate that the identity of this single nucleotide determines the binding specificity between guanine and adenine, which provides an example of how complex riboswitch structures could mutate to recognize new metabolite targets.

i. Phylogenetic Comparison Between Riboswitch Domains

A comparative sequence strategy was used to identify a series of intergenic regions from a number of prokaryotic species that carry a conserved sequence element termed the “G box” (see Example 6). B. subtilis carries at least five of these motifs, which were also identified using genetics techniques (Johansen, L. E., et al., J. Bacteriol. 185, 5200-5209). Each representative of the phylogeny has three potential base-paired elements (P1 through P3) and as many as 24 nucleotides that are conserved in greater than 90% of the examples identified to date. A subset of this phylogeny with features common to the G box motif highlighted is presented herein (Fig. 35A). When selected representatives are examined in greater detail, they are encompassed by the miRNA transcript of the gene immediately downstream, and thus are present as RNA elements located in the 5'-UTR of certain mRNAs.

Several notable differences present in the guanine-binding domain of xpt (Fig. 35B) relative to the RNA from ydhL (Fig. 35C) were identified. First, among the 23 sequence variations in ydhL, compared to xpt, 20 reside within base-paired elements and most of these changes permit base pairing to be retained. This strongly indicates that the overall secondary structure between the two RNAs is similar. Second, the remaining three mutations reside in unpaired regions, such that two (corresponding to positions 31 and 48 relative to xpt) reside at locations that are known to be variable. These mutations do not impact significantly the structure and function of the RNA. Third, the remaining mutation is a C to U change at position 74 relative to xpt, which otherwise corresponds to a strictly conserved nucleotide of the three-stem junction. Given the location of this mutation, this change might alter the molecular recognition characteristics of the ydhL aptamer.

Variant G Box RNAs Selectively Bind Adenine

It had been established (see Example 6) that the xpt aptamer makes numerous contacts with its ligand, and that as many as seven hydrogen bonds might be involved in forming the RNA-ligand complex. Furthermore, there is evidence that steric clashes also likely aid in restricting the range of metabolites that can be bound by the RNA. This array of contacts can only be established by forming multiple interactions between the various sides of guanine and distal parts of the RNA.

An intriguing hypothesis is the possibility that the C residue at position 74 of xpt could conceivably be forming a Watson-Crick base pair with guanine, thus forming three of these hydrogen bonds. Since a U mutation resides in the corresponding position in B. subtilis ydhL and two RNAs from C. perfringens and V. vulnificus, we believe that these RNAs might serve as adenine-responsive riboswitches. This hypothesis was further supported by recognition that the latter two genes (add) encode adenine deaminase enzymes. It seems reasonable that adenine should be the metabolite whose concentration is being monitored to determine the expression levels of adenine deaminase.

The ligand specificity of five G box RNAs (Fig. 35A) was examined by using in-line probing (Soukup, G. A. & Breaker, R. R. RNA 5, 1308-1325 (1999); Soukup, G. A., DeRosa, E. C., RNA 7, 524-536 (2001)). In this assay, the spontaneous cleavage of RNA is monitored in the absence of ligand, or in the presence of guanine or adenine. As predicted previously (see Example 6), the purE RNA (Fig. 36A) exhibits changes in the pattern of spontaneous cleavage products in the presence of guanine that correspond to that observed for the xpt RNA (Fig. 36B). These results confirm that the purE RNA, like the xpt RNA, responds allosterically to guanine and not to adenine when incubated in the presence of the concentrations of ligand tested.

In contrast, all three RNAs that carry the C to U mutation in the junction between P1 and P3 (corresponding to C74 of xpt) do not respond to guanine, but exhibit structural modulation only when incubated in the presence of adenine. Furthermore, the patterns of spontaneous cleavage for the adenine-specific aptamers are consistent with the secondary-structure model proposed for G box RNAs (Fig. 35). These results indicate that certain variants of the G box class of RNAs serve as sensors of adenine. Furthermore, these findings are consistent with the hypothesis that, when located in their natural settings, the ydhL RNA from B. subtilis and the two add RNAs from C. perfringens and V. vulnificus serve as adenine-specific riboswitches.
iii. The ydhL Aptamer Binds Adenine with High Affinity and Selectivity

Another characteristic of riboswitches is the aptamer domains exhibit tight binding for their corresponding target compound, and they discriminate against analogs, in some cases, by orders of magnitude in apparent $K_d$. For example, the guanine riboswitch from *B. subtilis* xpt exhibits an apparent $K_d$ for guanine of ~5 nM, but binds adenine with an apparent $K_d$ that is at least 100,000-fold poorer. In-line probing assays were used to determine the binding affinities of the *B. subtilis* 80 ydhL RNA for these two purines. As expected, the RNA exhibits progressively changing patterns of spontaneous RNA cleavage fragments in the presence of increasing concentrations of adenine (FIG. 37A), but the pattern remains unchanged with increasing guanine concentrations as high as 10 μM (see below).

The bands corresponding to spontaneous cleavage fragments that undergo change with increasing adenine concentrations were grouped into four sites and the extent of cleavage relative to the total RNA present were quantitated. This data was used to generate a plot (FIG. 37B) that provides an estimate of the apparent $K_d$ for ligand binding. In this instance, half-maximal decrease in spontaneous cleavage at sites 1, 2, 4, and the corresponding half-maximal increase in spontaneous cleavage at site 3 occurs when approximately 300 nM adenine is present in the in-line probing assay. Thus, the ydhL aptamer binds adenine with an apparent $K_d$ that is similar to those exhibited by other classes of riboswitches.

The molecular recognition characteristics of 80 ydhL were further examined by using the same in-line probing strategy with a variety of analogs. For example, a series of purine analogs that are close chemical variants to adenine exhibit measurable binding to the RNA (FIG. 38A). The ligands with measurable binding, 2,6-DAP, A and 2-AP, PMA (listed in order of decreasing affinity), are all close analogs of adenine. Furthermore, the relative affinities of the RNA for various ligands provide some indication of the contact points that the aptamer likely uses to establish molecular recognition (FIG. 38A, bottom right). This model is consistent with the finding that a series of purine analogs fail to exhibit measurable binding to the 80 ydhL RNA (FIG. 38B).

The collection of purines that are recognized by 80 ydhL indicate that only the Watson-Crick base-pairing face of the purine ligand is recognized differently by the ydhL aptamer compared to the xpt aptamer. For example, modification at the C8 position (8-chloroadenine) prevents ligand binding, which implies that a steric clash between certain purines and 80 ydhL as was observed for the xpt aptamer (see Example 6). Interestingly, the fact that 2,6-DAP, and not adenine, is the tightest-binding ligand provides insight into the similarities between the ydhL and xpt aptamers. This observation suggests that the 80 ydhL RNA retains at least one of the two hydrogen bond acceptor contacts that were proposed to exist in the xpt aptamer. Thus, the molecular recognition characteristics of these RNAs are consistent with the ydhL RNA differing in molecular recognition from xpt with a pattern that can be explained by a change from a Watson-Crick guanine-C base pair in xpt to a Watson-Crick adenine-U base pair in ydhL.

iv. Swapping Ligand Specificity of G Box RNAs by Molecular Engineering

The idea that the xpt and ydhL RNAs might be deriving their specificity for guanine or adenine by a Watson/Crick base pairing interaction was examined in greater detail by using a molecular engineering approach. A similar approach was used previously (Wilson, K. S. & von Hippel, P. H. *Proc. Natl. Acad. Sci. USA* 92, 8793-8797) to change the ligand-rescue specificity of an abasic hammerhead ribozyme construct from guanine to adenine. Both wild-type (93 xpt and 80 ydhL) and mutant (93 xpt C to U and 80 ydhL U to C) forms of G box aptamers were generated and tested for binding activity with guanine and adenine (FIG. 39). The mutations correspond to nucleotide position 74 relative to the xpt sequence (FIG. 35B), which is suspected to be the determinant of molecular discrimination between guanine and adenine.

As observed previously (see Example 6), the aptamer based on xpt exhibits structural modulation only when incubated in the presence of guanine, and is able to shift the distribution of tritiated guanine (but not adenine) in an equilibrium dialysis assay (FIG. 39A). However, the 93 xpt RNA that carries a single C to U mutation at position 74 no longer is responsive to guanine, but exhibits structural modulation and binding activity during equilibrium dialysis only in the presence of adenine (FIG. 39B). In contrast, the wild-type 80 ydhL RNA is specific for adenine (FIG. 39C), while the corresponding U to C mutation at this critical nucleotide position alters binding specificity to guanine (FIG. 39D). Therefore, the primary determinant of the base specificity of G box aptamers is the C or U residue that is present in the junction between stems P1 and P3, and that this base most likely forms a conventional Watson-Crick base pair with its target ligand.

v. Mechanism of Genetic Control by the ydhL Adenine Riboswitch from *B. subtilis*

In most instances, riboswitches control gene expression in prokaryotes by allosteric interconversion between alternate base-paired structures. For example, a TPP riboswitch from the thiM gene of *E. coli* makes use of alternate base pairing to sequester the Shine-Dalgarno sequence of the mRNA in the presence of ligand, presumably resulting in reduced translation initiation (see Example 2). In contrast, TPP riboswitches from *B. subtilis* harness ligand-binding events to alter base-pairing patterns and form intrinsic terminator stems that cause transcription elongation to abort (Gusarov, I. & Nucler, E. *Mol. Cell*. 4, 495-504 (1999); Mironov, A. S. et al. *Cell* 111, 747-756 (2002)). Similarly, metabolite-mediated formation of transcription terminator stems is a mechanism used by certain examples of riboswitches that respond to FMN (see Example 3 and 6), SAM (see Example 7), guanine (see Example 6), and lysine (see Example 5).

The UTR sequence of the ydhL riboswitch was examined to assess whether there is evidence of a transcription termination mechanism. Consistent with this possibility is the fact that the 5'-UTR of the ydhL mRNA can form a large hairpin, composed of as many as 22 base pairs, followed by a run of eight uridyl residues (FIG. 40A). This structural feature, which was also noted elsewhere recently (Johansen, L. E., et al., *J. Bacteriol*. 185, 5200-5209), is characteristic of an intrinsic terminator stem. In the absence of adenine, it was considered that the riboswitch can form this intrinsic terminator. If true, then the genetic control status for this riboswitch would default to this predicted OFF state, which prevents gene expression by inducing transcription termination. In the presence of adenine, gene expression is expected to proceed because a substantial portion of the left shoulder of the terminator stem would be required to form stems P1 and P3 of
the adenine aptamer domain. Since stems P1 and P2 are integral components of the adenine aptamer, ligand binding would establish a structure that precludes formation of the terminator stem.

[0547] This mechanism for the ydhL riboswitch was assessed in vivo by generating reporter constructs wherein various forms of guanine- and adenine-specific riboswitches were integrated into the B. subtilis genome. As controls, two reporter constructs were prepared with either the wild-type xpt riboswitch, or the xpt variant with the C to U mutation at position 74. As expected, the wild-type xpt construct causes repression of β-galactosidase expression when presented with excess guanine in the culture medium (FIG. 40b). This finding is similar to those reported previously for function of the guanine riboswitch from xpt (see Example 6). Adenine also shows a modest (~4 fold) repression of reporter expression after a six-hour incubation. This latter effect is most likely due to the function of the PurR protein, which is known to provide modest down-regulation of transcription initiation in response to adenine at the xpt-phiX promoter used in this construct (Cristiansen, L. C., et al., J. Bacteriol. 179, 2540-1550 (1997)).

[0548] A near identical xpt construct carrying the C to U mutation causes a loss of regulation upon addition of guanine, but shows no change in the putative protein-dependent control due to adenine (FIG. 40c). However, these results are surprisingly consistent with the observed loss of guanine binding in vitro when this mutation is made, but suggest that the resulting specificity change to adenine in vitro does not persist robust adenine-dependent genetic control in vivo. Most likely, the diminished expression upon addition of adenine again is due to the PurR protein.

[0549] In contrast to the xpt riboswitch, the performance of the corresponding wild-type and mutant ydhL reporter constructs indicates that the latter is an adenine-depentent riboswitch with the opposite response to rising levels of ligand. Specifically, the wild-type ydhL construct exhibits very low β-galactosidase activity when assayed in the absence of ligand, or in the presence of guanine (FIG. 40d). However, a greater than 10-fold increase in gene expression occurs in response to added adenine. In addition, the single U to C mutation at the P1-P3 junction of the aptamer causes substantial (~100 fold) derepression regardless of what ligand is added (FIG. 40e). Although this seems counter to the model proposed for ydhL riboswitch function, it is important to note that this mutation indeed disrupts adenine binding, but it also causes a mismatch to occur in the terminator stem. If this mismatch is sufficiently destabilizing to the terminator stem, or if this mutation adversely affects the folding pathway for the riboswitch, then the default ‘OFF’ status for the genetic control element would be expected to change to default ‘ON’. Therefore, the observed level of gene expression might be indicative of full activation of the ydhL gene when its genetic control element is indifferent to the concentrations of purines in the cell.

[0550] 2. Discussion

[0551] i. The Structure and Evolution of Adenine Riboswitches

[0552] The sequence and biochemical similarities between guanine- and adenine-specific G box RNAs indicate that they are analogous in overall secondary and tertiary structure. The ease of interchanging ligand specificities of these aptamers by making single mutations to the xpt and ydhL aptamers suggests that such changes might occur with high frequency in natural populations. However, the fact that neither single-base variant of the xpt or ydhL riboswitches exhibits corresponding specificity changes in genetic control in vivo suggests that multiple mutations might be necessary to make a usefulswap in riboswitch specificity.

[0553] It is important to note that the binding affinity of the resulting single-base xpt variant is not as robust for its new ligand. Specifically, the wild-type xpt RNA has an apparent Kd for guanine of no greater than 5 nM (FIG. 39a), while the C to U variant of this RNA exhibits an apparent Kd for adenine of ~100 nM (FIG. 39b). In this case, although the mutation results in a substantial change in base discrimination between guanine and adenine, binding affinity for the matched ligand has been somewhat degraded. In contrast, the wild-type and mutant ydhL RNAs exhibit both specificity change and retention of binding affinity for the matched ligands (FIGS. 39c and 39d). However, the affinity for the U to C variant of 80 ydhL for guanine appears to be at least 10-fold poorer than that of 93 xpt.

[0554] Thus, accessory mutations that do not directly define ligand specificity but that further adjust the binding affinity might be necessary for G box RNAs to interconvert between guanine and adenine ligands in a biological setting. In this regard, it is interesting that the ydhL and xpt aptamers differ from each other at 23 positions (FIG. 35), with only one residing within an obviously critical position (C74 of xpt). Although some of these mutations might serve to fine-tune the binding affinity of the aptamers, many could be the result of neutral drift in the RNA sequence that is permitted because they retain the essential secondary-structure elements.

[0555] ii. Genetic Control and Function of the ydhL mRNA

[0556] Mutant strains of B. subtilis that resist the toxic effects of 2-fluoro-adenine were reported recently (Johansen, L. E., et al., J. Bacteriol. 185, 5200-5209)). These mutations, which cause over-expression of the ydhL gene product, were mapped to the adenine riboswitch domain. In both instances, the changes (deletions) are expected to disrupt riboswitch function by eliminating a portion of the terminator stem or by eliminating both the terminator stem and portions of the adenine aptamer domain. In both instances, the variants preclude the riboswitch from adopting its default state (transcription termination), which causes unmodulated activation of gene expression.

[0557] The protein product of the ydhL gene (also termed pbuU) has been proposed to be a purine efflux pump (Johansen, L. E., et al., J. Bacteriol. 185, 5200-5209). Thus the resistance to 2-fluoro-adenine conferred upon the cell by disruption of the adenine riboswitch from ydhL might be due to excretion of this toxic compound. In the natural genetic background, the presence of excess adenine within the cell most likely induces increased expression of the ydhL gene to produce the purine efflux protein. Higher levels of this protein then work to normalize the concentration of purines by pumping out of the cell one or more forms of this compound class.

[0558] iii. Riboswitch Mechanisms—Genetic Activation and Deactivation by Rising Metabolite Concentrations

[0559] The adenine riboswitch from B. subtilis also is notable for its mechanism of action. In the majority of riboswitches examined to date, metabolite binding causes a lowering of gene expression. This occurs either by ligand-mediated formation of a terminator stem to prevent transcription of the complete mRNA, or by sequestering the Shine-Dalgarno sequence and precluding translation initiation. In most instances, the down-regulation of gene expression is
expected, as a build-up of sufficient levels of a particular metabolite should logically provide a signal to turn off genes that are involved in biosynthesis or import of the compound (Grundy, F. J. & Henkin, T. M. et al., *Frontiers Biosci.*, 8, D20-31 (2003)).

**0560** The adenine riboswitch from ydhL (and presumably for the add riboswitches as well) belong to a group of genes whose functions would hint at the need for riboswitch activation in the presence of high concentrations of target compounds. In the case of ydhL, disposal of excess purines would seem to be an important capability given that certain purines such as guanine are insoluble at modest concentrations. Alternatively, there be no obvious need to express adenine deaminase if adenine concentrations were exceptionally low, and therefore we expect that the riboswitches from the add genes of *C. perfingens* and *V. vulnificus* might be activated by ligand binding as well. Interestingly, T box domains, which are 5'-UTR structures that control the expression of many aminoacyl-tRNA synthetases in *B. subtilis* and other Gram-positive organisms (Grundy, F. J., et al., *Proc. Natl. Acad. Sci. USA* 99, 11121-11126), also induce gene expression in response to raising concentrations of the target they sense. However, unlike the known metabolite-binding riboswitches, T box domains sense the biochemical precursor (non-aminoclayl tRNAs) to the products of the enzymes whose expression they control (Miller, J. H. A Short Course in Bacterial Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1992)).

**0561** Although we expect that riboswitches that induce gene activation in response to increasing metabolite will occur less frequently due to genetic necessity, there is no inherent structural flaws in RNA folding that would skew this distribution between gene-activating and gene-deactivating riboswitches. Whether the riboswitch responds to ligand binding by activating or repressing gene expression, the RNAs will exploit allosteric changes in secondary and/or tertiary structure that are based on the same principles of RNA folding. The only obligate difference between activating and repressing riboswitches is in the fine structure of the expression platform, whereas the aptamer domain can remain largely unchanged.

**0562** 3. Methods

**0563** i. Purine Analogs

**0564** Guanine, adenine, 2,6-diaminopurine, 2-aminopurine, hypoxanthine, xanthine, 1-methyladenine, purine, 6-methylaminopurine, N9-N9 dimethyladenine, 6-mercaptopurine, 3-methyladenine, guanine-8-1H and adenine-2,8-2H were purchased from Sigma. 6-cyanopurine and 8-azadename were obtained from Aldrich and 2-chloroadenine, 8-chloroadenine from Biolog Life Science Institute, Germany.

**0565** ii. DNA Oligonucleotides

**0566** Oligonucleotides were synthesized by the HHMI Keck Foundation Biotechnology Resource Center at Yale University, purified by denaturing polyacrylamide gel electrophoresis, and were eluted from the gel by crush-soaking in a buffer containing 10 mM Tris-HCl (pH 7.5 at 25° C.), 200 mM NaCl, and 1 mM EDTA. DNAs were precipitation with ethanol, resuspended in deionized water, and stored at -20° C. until use.

**0567** iii. In-Line Probing of RNA Constructs

**0568** RNA constructs were synthesized from the corresponding PCR DNA templates by transcription in vitro using T7 RNA polymerase, dephosphorylated, and 5'-end labeled with 32P as described in Example 6. In a typical in-line probing assay, 2 nM of labeled RNA were incubated in a buffer containing 20 mM MgCl2, 50 mM Tris-HCl (pH 8.3 at 25° C.) and 100 mM KCl in the absence or presence of purine compounds as indicated for each experiment for 40 hrs at 25° C. Purine concentrations ranging from 1 mM to 10 μM were employed unless otherwise noted. At the end of each incubation, spontaneously cleaved products were separated on a denaturing (8 M urea) 10% PAGE, visualized using a PhosphorImager and quantitated using ImageQuaNT software (Molecular Dynamics).

**0569** iv. Equilibrium Dialysis

**0570** Equilibrium dialysis assays were conducted using a DispoEquilibrium Dialyzer (Harvard Biosciences), wherein chamber A and B are separated by a 5000 MWCO membrane. Chamber A contained 30 μl of 1H-guanine or 1H-adenine at a concentration of 100 mM in a buffer containing 50 mM Tris-HCl (pH 8.5 at 25° C.), 20 mM MgCl2, and 100 mM KCl. A 30 μl aliquot of the above mentioned buffer containing RNA at 3 μM concentration was delivered into chamber B. Equilibrations were allowed to proceed for 10 hrs at 25° C. Subsequently 5 μl was withdrawn from each chamber and quantitated by liquid scintillation counting.

**0571** v. Construction of xpt- and ydhL-lacZ Fusions

**0572** A DNA construct encompassing nucleotides 468 to +9 relative to translational start site of ydhL was PCR amplified from *B. subtilis* strain 1A40 (Baci1tus Genetic Stock Center, Columbus, Ohio) with primers that introduced EcoRI-BamHI restriction sites. The wild-type construct was cloned into pDG1661 at EcoRI-BamHI restriction sites directly upstream of the lacZ reporter gene and sequenced to confirm its integrity. The resulting plasmid was used as a template for site-directed mutagenesis via the QuickChange site-directed mutagenesis kit (Stratagene) using the appropriate primer. Plasmid variants were integrated into the amyE locus of *B. subtilis* strain 1A40 and the transformants were confirmed as described in Example 6.

**0573** vi. In Vivo Analysis of Riboswitch Function

**0574** Transformed *B. subtilis* cells were grown to mid log phase with constant shaking at 37° C. in minimal media containing 0.4% w/v glucose, 20 g/l (NH4)2SO4, 25 g/l K2HPO4, 6 g/l KH2PO4, 1 g/l sodium citrate, 0.2 g/L MgSO4, 7H2O, 0.2% glutamate, 5 μg/ml chloramphenicol, 50 μg/ml L-tryptophan, 50 μg/ml L-lysine and 50 μg/ml L-methionine. Guanine or adenine was added to a final concentration of 0.1 mg/ml. Cells at mid exponential stage were harvested and resuspended in minimal media in the presence or absence of purines and grown for an additional time as indicated for each experiment, at which time 1 ml of cell culture was subjected to β-galactosidase activity assays using a variation of the method described by Miller (Miller, J. H. A Short Course in Bacterial Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1992)).

I. Example 9

Tables of Sequence Comparisons for the Sam, Cobalamin, Guanine, Adenine, and Lysine Riboswitches Discussed Herein

**0575** FIG. 41 shows sequence and types of riboswitches. The alignment of these sequences is as disclosed herein, regions disclosed in the other figures correspond to the same regions in FIG. 41.
Additional riboswitches were found based on published alignments and secondary structures (Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* 30, 737-749 (1998)) using the SequenceSniffer program. This program finds degenerate matches to RNA patterns defined by linked sequence motifs and base pairing constraints. In the alignments, base pairing regions have the identical underline styles or boxes and are labeled as in the corresponding figures discussed in Examples 1-8, with the addition of a putative pseudoknot marked PS. Predicted terminators (short dashed underline) and start codons (long dashed underline) are marked for some sequences. Positions for each sequence in the indicated Genbank record or unfinished genome contig are for the sequence column marked with a circle (●)—the fifth base in stem P1 that is 5' of the aptamer. Start is the offset from the column marked with an asterisk (*)—the sixth base in stem P1 that is 3' of the aptamer—to the start codon of the first gene in the operon. Genes were identified from COGNIITOR (Tatusov, R. L., et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22-28 (2001)) and PFAM (Bateman, A., et al. The Pfam Protein Families Database. *Nucleic Acids Res.* 30, 276-280 (2002)) database matches to protein sequences annotated in the Genbank records. The standard names from these databases are used when possible (2011-COG2011; ???-no matches). Previous operon designations for *B. subtilis* are given in parentheses (Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. *Mol. Microbiol.* 30, 737-749 (1998)). A subset of sequences with <90% pairwise identity between the bases encompassed by stem P1 was selected for determining the consensus sequence. In the consensus sequence, lowercase and uppercase bases indicate >80% and >95% conservation at a position, respectively. Purine (R) and pyrimidine (Y) bases were assigned when no single base had >80% conservation.

(*) Sequence shares >90% identity with another sequence, and was excluded when determining the consensus.

(1) Very short hypothetical gene that may be a misannotated ORF.

(2) Possible S Box "pseudogene". The S Box is on the opposite strand 5' of the indicated operon.

It is understood that the disclosed method and compositions are not limited to the particular methodology, protocols, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a riboswitch" includes a plurality of such riboswitches, reference to "the riboswitch" is a reference to one or more riboswitches and equivalents thereof known to those skilled in the art, and so forth.

"Optional" or "optionally" means that the subsequently described event, circumstance, or material may or may not occur or be present, and that the description includes instances where the event, circumstance, or material occurs or is present and instances where it does not occur or is not present.

Ranges may be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, also specifically contemplated and considered disclosed is the range from the one particular value and/or to the other particular value unless the context specifically indicates otherwise. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another, specifically contemplated embodiment that should be considered disclosed unless the context specifically indicates otherwise. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint unless the context specifically indicates otherwise. Finally, it should be understood that all of the individual values and sub-ranges of values contained within an explicitly disclosed range are also specifically contemplated and should be considered disclosed unless the context specifically indicates otherwise. The foregoing applies regardless of whether in particular cases some or all of these embodiments are explicitly disclosed.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of skill in the art to which the disclosed method and compositions belong. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present method and compositions, the particularly useful methods, devices, and materials are as described. Publications cited herein and the material for which they are cited are hereby specifically incorporated by reference. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such disclosure by virtue of prior invention. No admission is made that any reference constitutes prior art. The discussion of references states what their authors assert, and applicants reserve the right to challenge the accuracy and pertinency of the cited documents. It will be clearly understood that, although a number of publications are referred to herein, such reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

Throughout the description and claims of this specification, the word “comprise” and variations of the word, such as “comprising” and “comprises,” means “including but not limited to,” and is not intended to exclude, for example, other additives, components, integers or steps.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the method and compositions described herein. Such equivalents are intended to be encompassed by the following claims.
SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 4

<210> SEQ ID NO 1
<211> LENGTH: 202
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 1

gccggucucc ugaaauaagg gagaucucugu gacucgcgccc gcgcgguaa 60
ggaaggagacu gauagucacuc gcgcggccc gcgcggggc gggggccc gaucaucuc 120
uauauuaagg caaaggccggaa gacucgcgc gcgcggccgc gcgcggcgcu gauccucu 180
caucaucgc uauaauugga 202

<210> SEQ ID NO 2
<211> LENGTH: 165
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 2

<210> SEQ ID NO 3
<211> LENGTH: 240
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 3

<210> SEQ ID NO 4
<211> LENGTH: 165
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 4
<400> SEQUENCE: 4

ggaacacaag caccaggggu gcccwacgucg gugwggcgcg agsaauaccc guauccagug 60
aucwagaua ugcwagcuag ggsaagucac ggacaccagg guauwsecu cuucaugua 120
ugccaggacg wacuauagca aguagaccuc cuggauccag cgcaaa 165

<210> SEQ ID NO 5
<211> LENGTH: 176
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39-156
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 5

ggaauauagc cgusaggguc gaagcgcacc cugausgmn nnnnncaga gaagcaggg 60
gacugcccq acagacccgc ggcaacccgg guauugggra ucaugacaua ccagggcgu 120
asuacccgca agcucaaca gcuugggaag nnnnnccga acguagcgc gagcuc 176

<210> SEQ ID NO 6
<211> LENGTH: 4
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 4
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 6

ggun

<210> SEQ ID NO 7
<211> LENGTH: 6
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 6
<223> OTHER INFORMATION: d = g, a or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-4
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 7

nnnngd

<210> SEQ ID NO 8
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 11, 17, 20, 25, 36
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 6, 36
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 13, 15, 31
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 8

yyucrggsc nggynaan ucctnaccg yggurn

36

<210> SEQ ID NO 9
<211> LENGTH: 51
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 7-9, 13, 14, 16, 18, 25, 26, 32, 33, 37, 39, 42, 43, 50, 51
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 38, 44
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 17, 34
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 9

ncuauung agugnynaga ggagnggcc cnnyannc cnnygcacn n

51

<210> SEQ ID NO 10
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 2, 10-17, 22, 25-31, 34, 40-46, 54-60, 69, 69
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 5, 16, 67
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 65
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 10

nnucruaau nnmnnnnnau anggnnnnnn ngunucuacn nnmnncngu aaannnnnnn

60

acwuyrnn

69

<210> SEQ ID NO 11
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note =
synthetic construct

<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 2, 10-17, 22, 25-31, 34, 40-46, 54-60, 68, 69
<223> OTHER INFORMATION: n = g, a, c or u
<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 5, 16, 67
<223> OTHER INFORMATION: r = a or g
<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 65
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 11

nnucrumauan nnmrrrrrrr rrrrrrrrrr rrrrrrrrrr rrrrrrrrrr

<400> SEQUENCE: 69

auuaygrnn

<210> SEQ ID NO 12
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FRAGMENT:

<223> OTHER INFORMATION: Description of Artificial Sequence: /Note = synthetic construct

<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 13-16, 20, 21, 26-33
<223> OTHER INFORMATION: n = g, a, c or u
<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 2, 12
<223> OTHER INFORMATION: r = a or g
<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 3
<223> OTHER INFORMATION: w = a or u
<220> FRAGMENT:

<221> NAME/KEY: misc_feature
<222> LOCATION: 8
<223> OTHER INFORMATION: h = a or c or u

<400> SEQUENCE: 12

rwaasaghgo rrrrrrrrrrrr aauunnnnn nn

<400> SEQUENCE: 33

<210> SEQ ID NO 13
<211> LENGTH: 165
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 13

ggagaccaau aguuaauucc uuggggcagg gugsaauucc cagcggcagg 60
uaguaasgc cauucugcuu uggcggcgg caccgcggg accagccgc ccggcgcgg 120
guuaaaguc gcagcaggcag uggcagcgcug gugggcgg cgg 165

<210> SEQ ID NO 14
<211> LENGTH: 128
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana

<400> SEQUENCE: 14

gggcagatc gagccacac caccaggcgg gcuugaacca gggcagccgc cgcagaagcc 60
gggcagacgc gcagcggcag uggcagcgc uggcagcgc uggcagcgc uggcagcgc 120

gggcgucct 128
<210> SEQ ID NO 15
<211> LENGTH: 135
<212> TYPE: DNA
<213> ORGANISM: Oryza sativa
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 33-83
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 15

ggugaaugacagcaagcaccaggggu gennyynynyn nnynynnynnn nnnynnnnn 60
nnynynynnn nnynynynnn nnynynynnn nnnynnnnn nnnynnnnn 120
cucguagagc agugu

<210> SEQ ID NO 16
<211> LENGTH: 135
<212> TYPE: DNA
<213> ORGANISM: Poa secunda
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 33-83
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 16

ggugaaugacagcaagcaccaggggu gennyynynyn nnynynynnn nnnynnnnn 60
nnynynynnn nnynynynnn nnynynynnn nnnynnnnn nnnynnnnn 120
cucguagagc agugu

<210> SEQ ID NO 17
<211> LENGTH: 176
<212> TYPE: DNA
<213> ORGANISM: Neurospora crassa
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 15-123
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 17

gcuacccgggu guccynynyn nnnynynynnn nnnynnnnn nnnynnnnn nnnynnnnn 60
nnynynynnn nnynynynnn nnynynynnn nnnynnnnn nnnynnnnn nnnynnnnn nnnynnnnn 120
nnnggucuga gaaaucccg gcacucugau cuggaaau ccacggaag gauggc 176

<210> SEQ ID NO 18
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 9
<223> OTHER INFORMATION: d = g, a or u
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-7, 10-16
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 18

nnnnnnngdn nnnnnncuga ga 22

<210> SEQ ID NO 19
<211> LENGTH: 103
accaacgag uncggggugm nnmnmnmn nnmncugag amnmmmmmn naauaccgg

auuccaguc uncggggugm ccgoggaacc gauggcccm accgoggaacc

uggacaccgg aauaccucgc aagggacaa gaguuaa

auauuucug uncggggugm uuaugacccu uauucggggu ccgoggaacc

uggacaccgg aauaccucgc aagggacaa gaguuaa

auauuuagc uncggggugm uuaugacccu uuuuugacccu uauucggggu ccgoggaacc

uggacaccgg aauaccucgc aagggacaa gaguuaa

caacacagc accgggagm nnmnmnmn mnmmnmnmn nnmnmnmn mnmmnmnmn nnmnmnmn

nnmmnmnmn mnmmnmnmn mnmmnmnmn nnmnmnmn mnmmnmnmn

nnmmnmnmn mnmmnmnmn mnmmnmnmn mnmmnmnmn mnmmnmnmn

nnmmnmnmn mnmmnmnmn mnmmnmnmn mnmmnmnmn mnmmnmnmn

cacgcgaacc auuuggcuucu ugu
<400> SEQUENCE: 23
"cuuagcgug gacgcaagct nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nncugagann nnnnnnnnu aaacggucuaa acuuguacucu gauauaccuac ccagaaacuggu 180
cauagcucu 190"

<210> SEQ ID NO 24
<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Fusarium oxysporum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12-117
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 24
"auucaugcaug anccggggun nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnn nnnnnnnnn nncugagann nnnnnnnnu aaacggucuaa acuuguacucu 120
gauauaccuac ccagaaacuggu 156"

<210> SEQ ID NO 25
<211> LENGTH: 156
<212> TYPE: DNA
<213> ORGANISM: Fusarium solani
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12-113
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 25
"auucaugcaug anccggggun nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnn nnnnnnnnn nnnnnnncuc gcgannnnnn nnnnnnuaacgc gcggaccucu 120
gauuagguau aaccacgcu aacgauacgc cuccuc 156"

<210> SEQ ID NO 26
<211> LENGTH: 133
<212> TYPE: DNA
<213> ORGANISM: Arabidopsis thaliana
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12-81
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 26
"gcaaaag cac naaaggggun nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccuagcuuc nnnnnnnaaaccucn aaccgguaa acggggaacg ccugugcagcggg 120
gauagcucu uuu 133"

<210> SEQ ID NO 27
<211> LENGTH: 140
<212> TYPE: RNA
<213> ORGANISM: Poa secunda
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12-88
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 27
-continued

aaagugcaac.cnaggggugn.nnnnnnnn.nn.nn.nn.nnnnnnn.nn.nn.nn nn.nn.nn.nn.nnnnnnn 60
nnnnnnnn. nncugagann.nnnnnnnn.nngcuwuugga.nncagacag. gauauggcu 120
ngugaagggg. uugucauuc 140

<210> SEQ ID NO 28
<211> LENGTH: 140
<212> TYPE: DNA
<213> ORGANISM: Oryza sativa
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12-86
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 28

aaagugcaac.cnaggggugn.nnnnnnnn.nn.nn.nn.nnnnnnn.nn.nn.nn.nn.nnnnnnn 60
nnnnnnnn. nncugagann.nnnnnnnn.nngcuwuugga.nncagacag. gauauggcu 120
ngugaagggg. uugucauuc 140

<210> SEQ ID NO 29
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 26-190
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 29
cygugagguu. gagugugcag.udauauaagn. agauuaucau. uucugnnngn. agnnaagug 60
nnnnauagau. gaaaggaaug. uungaagga. aguauuaguc. ogaaauaggu. uuguguccaco 120
auuungcaca. cuuugcugugu. cuuguauuaa. uauangugca. angncuguc. acaacgguuu 180
nnnnnnnn.nnguaugug. gacauugc. gagg 214

<210> SEQ ID NO 30
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 25-191
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 30
cuuaaagguu. gagcogcogca. uaggnnaag. aguauuacu. auggnnnngn. agnnaaagug 60
nnnnnaannn. nnnnnngu. uungaaggg. acuaunnuuc. ogaaauuaa. gauaaccau 120
nnncuaauca. uauuugccua. cuuguauuun. gaaauaag. aguauucguc. auugauuaa 180
nnnnnnnn.nnuuauugga. gacacuuug. gaga 214

<210> SEQ ID NO 31
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 26-165
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 31
cgaugagguua gagguggcga cuuuunaagnn aguannaaac ggacnnnnngn agauacgaga
amnnynugcu aaguaccnggu uunngaagga aaegnnnguc ggaaguuuaau auuuccuucuc
unnngaaauua uagacguggg cugugunnnu gaangaagaa aagacucugac aeguuuaca
aaauaacccug uaaaccguggg guguaccuccu acgc

<210> SEQ ID NO 32
<211> LENGTH: 214
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-189
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 32
agugagguua gaggungcga aaaccaagn aguannaca aauunnnggn agnnngaaau
60
gagacnccgu aguacaguauu gnngaaagg gaannnuuuc gcgaagcugg aagacucacau
120
nnngacccug agucnccggcu cguuauunnn aauaaacsac aagacucugac aauauacgga
180
uguuunnnnu gcuauaggg gggguaccuc acgc
214

<210> SEQ ID NO 33
<211> LENGTH: 214
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-187
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 33
agugagguua gaggungcga aaaccaagn aguacnacac ucnnnguagn agnnaauggac
60
aaucngucac nnunggcacu uuggaaggg ggnnnucuug cgaagucgac auuggggcu
120
aunnucuacu agucgcugc acuullnunnn guuuuccuac ugggucuguc acacacacug
180
ccuaacuaac gacugcugac gacuauaccu acgc
214

<210> SEQ ID NO 34
<211> LENGTH: 214
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-191
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 34
agugagguua gaggangcgg guuunuaagn aguacnccgu uuggnnnnng aggaacacaa
60
nnnnncggag amnnuaagcg cncgaaagga aamnucucgc ggaagcggaa gaaagucucac
120
nnnncuggc cuuggcuggg uuuguunnn guuuaaugga aacuacugcc aacucgcau
180
nnnnnnnnn nuugugcugac gacuaccuccu cuuu
214

<210> SEQ ID NO 35
<211> LENGTH: 214
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: mioc_feature
<222> LOCATION: 16-191
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 35

ggugaaagau gaggungcga acuucnaagn aguauungccu uunnnnnggn agnaaagaug 60
ganuunuc guguaggcngu gggcngggc gagaaccaau aaaaaccccaau 120
cnnggguuuu uuugccccg gugcawumnn gaauaauggu aaggnugcugc aagaaaucau 180
nnnannnnnn nuuucuugg gggcuaucuc guug 214

<210> SEQ ID NO 36
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-165
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 36

acccuuuugga gaggungcgu uaaacaagc aguuaccggg uuggngnnn agnnuuucgca 60
nnnnnaacu agauuagacg gnuuacggg guuucnnacg caagacauuu gauuuggcan 120
nnngauuuu uuucuucggcu uucuauanmn caucauaaga augncugucu acuuccuauag 180
uuuuuauu uguauuggu ggggacaag gcuacacagu gac 214

<210> SEQ ID NO 37
<211> LENGTH: 215
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-193
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 37

gacccaaagau gaggugcgg gauuauaagc aguanuuacg uumnnnangn agngcugaca 60
nnnnnnnngu unnmmnuag auunngaga gauunnuagg caaagagagau uuuuauggg 120
nnnnnaauaa uuuuucuuggu uaauguauun nnaunegucg uuuuacgauu ccuuccuuas 180
nnnnnnnnnn nnaaagugug agucuacaa guao 215

<210> SEQ ID NO 38
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 38

aaccaguau gaggungcga uauuunaan aguanucug uugnnmmagn agnguuaaagc 60
nnnnauaua anuuccaaag guguaagcga agaaccuauu agaaagggcu 120
uuauauauu uuggguuggg uguauauun gaauuuuugga aacuacucugc aucuauaun 180
nnnnnnnnnn nnuuucugug gguuacuau caaa 214

<210> SEQ ID NO 39
<211> LENGTH: 214
<210> SEQ ID NO 40
<211> LENGTH: 225
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-194
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 39

aagagggua gaggngcga gsaucaaggn auaanncaa aaunnnngg angnnuuaag

nnnnagccu agaaguuua gnrngaagg gauaunncgc cgaaguuuuu ggcuaauacu

uuanggcuu aaugcuggg uguauajnnn gauaauacac acacuguc acaaajnnn

nnnnnnnnnnnnnnnagggcagcuaacu cuaa

214

<210> SEQ ID NO 41
<211> LENGTH: 224
<212> TYPE: DNA
<213> ORGANISM: E.coli
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-204
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 40

cagcgcagaa gaggngcgu ugcoccnann aguaacgguug uagwnnnnngn agngcaca

gnnnnucug ugsuanuacae cnnnnnugg ggggccucgc ggagcggauu gaacggcugg

ccacngcuu aucacgguug acaggggncu gauucccccu gggnnnguc gccannnnnn

nnnnnnnnnnnnnnnnnagggcagcugacug ggugu

225

<210> SEQ ID NO 42
<211> LENGTH: 215
<212> TYPE: DNA
<213> ORGANISM: Haemophilus influenzae
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-191
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 41

uacaaagga gagcngcgsa uaaunsean agrnmuuag uwnnnnnnagn aggnnggaas

ammngcag qanngaaass aarngaaaga aauagngucog quaauaaa uaaaaacgn

nnnnnuuugg uggcugugg gcggcucmnn gaaanggggc gacacacuguc aauguuuuc

ugauunnnnaaacagggguu guggcagcgg uguu

214

<210> SEQ ID NO 43
<211> LENGTH: 216
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 42

guuuugggua gaggngcgg gcaggcnacuq aguannuaa agcnnngnga agnggsaaau

gagncncnnn nnnnngcua ugnngaaaggg ggnnnucug ccgaagcggc ggaauauacuc

auuacacuun acucugcgg ggucacuun ngacacaaac acacucugu cauaacgcg

annnnnnnnnnnnnncuaauugg agggcuacgg acgcg

215
<210> SEQ ID NO 43
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 43

ucgguggguagaggacu ccaacnaaun aguannauag acnnaagn aggauggacaa
nnnnngaug aaannguugg uunnggaagg uugguunuug cgaacnaaau uuagggucag
annnuuaauu auucguggua caacuunnn naaauaaaga ugoancuguc auuoaaatu
aagmnannnn nnncaaugga gaacuacuga ucga

<210> SEQ ID NO 44
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Pasteurella multocida
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 44

uacuuggaugu gagaggcga ucacuaan auuannnuuu uucnnmmng agnnuggaua
annnngeag arngaaaaga grngaaagga guaannmngc caaauucaau ugaagaucan
nnnnnuuugg uuugguuggug cgguaauuun nnnanggaac guanuuuguc auugucuuuu
uuaaannnnn nnacuacugu gogcuacugg uugg

<210> SEQ ID NO 45
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-191
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 45

auuauuugau gaggngcau ccaucauagn aguannauag uuanannngn auuuaucug
cuuuuuuucu cuuuuauuu uunnguaggg uunguagguc cgaacgauu auuaagcan
nnnnnuuaau uuuuguuggac uuuuugunn uuagacguga gguunnuguc auuuuuuu
nnnnnnnn naaauaugga gugcaaucu uguu

<210> SEQ ID NO 46
<211> LENGTH: 216
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 26-196
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 46

auuuggaugu gaggugcau guuuaanuun aguannacuu guunnnccaga agnnuuuuu
ugnnnuunnn nnnnnncau agungaag guaagnnauu gcggaaauag auuaaccu
uaanannnuua uaucuauugg gacaguuzuun ngcaagagga acuguacucg ucacagaann
nnnnnnnnnnnnnnnnnug augugcuacc uuaauu

<210> SEQ ID NO 47
<211> LENGTH: 216
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 47
agauuuuuagagggcaggucaugaugnuug ngcaguguuau ucacauua cagacauuau
uugguacacc uuuuuguguu uggauuug uguuuuuuuu

<210> SEQ ID NO 48
<211> LENGTH: 216
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 26-196
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 48
auaaagagug auucuagccau uuuuuuua cagacauuau uaaauuagc auuauuuc
uugguagcagggcaggucaugaugnuug ngcaguguuau ucacauua cagacauuau
uugguacacc uuuuuguguu uggauuug uguuuuuuuu

<210> SEQ ID NO 49
<211> LENGTH: 225
<212> TYPE: DNA
<213> ORGANISM: Shigella flexneri
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-204
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 49
cacgcuagcag gacgagcag uacgacnuuu ngcaguggu uugnnnnnngngcagagca
uugguacacc uuuuuguguu uggauuug uguuuuuuuu

<210> SEQ ID NO 50
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Shevanella oneidensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-194
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 50
ccannuuccu ucacuacac ccagnuugg ggacacacuc ccagacauuau gaagcgggguu
uugguacacc uuuuuguguu uggauuug uguuuuuuuu
aggaacagaa gaggangcgu uuaecunnmm ngguannuccu auaungaggm acacaaacuc 60
cgcgcaummm ngmuugaa saanggaggga guuuwagcqc cgagcaaug auguggcuc 120
ugnucaaugu uauguggcgu gcuuagcncu gauncocuca ccgunuguc accuuaaau
nmnmnmnm nmngguugga ggcuucuccg ugcac 214

<210> SEQ ID NO 51
<211> LENGTH: 214
<212> TYPE: DNA
<213> ORGANISM: Shevanella oncidens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-192
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 51
ccuuaaagcu gagccgagcg ugcounuaugn auuanncucug uugognmnmn agauggcguau
60
gnnncocga gunaaguacca gngaaagga guuncaagcgc cgaauagucg accuacuacca
120
nnnnnnnnn agccgguugn ugcounuacaa auuanncucua agaugugucc accuacucc
180
nnnnnnnnn nncuacuagc ggcuucuccu augg 214

<210> SEQ ID NO 52
<211> LENGTH: 218
<212> TYPE: DNA
<213> ORGANISM: Thermatoga maritima
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-194
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 52
ugacccgacg gagccgagcg cccaghagug accuagnncnn ngnnagggau
60
cgnnnnnnnn nnnnngggaa ccgguuacca agggaggg qccgoacaggg guacacaggu
120
cucuocggc cucaugucgu ggggguugg gnnnacuaccc caucuccnnncu uggcuacgqag
180
guucnnnnn nnnncuccac ugggacgcg aucuggcuc 218

<210> SEQ ID NO 53
<211> LENGTH: 215
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 16-188
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 53
agguagagcu ggcgngcgg gucaacagn auuannacau gcccnnnmgn agnnnggnuua
60
nnnnnngnn nnnnnguggu uguuugnnc qgggggnnc qgcgggaqcgy acuacuacuc
120
cuaacgguu uuacgacugcc gucaacgcn nngacacguu uagacgncg ucuacucqgg
180
ccuocguccg ccoccuacgug agacuacuc qgcuc 215

<210> SEQ ID NO 54
<211> LENGTH: 218
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc.feature
-continued

212 TYPE: RNA
213 ORGANISM: Vibrio vulnificus
220 FEATURE:
221 NAME/KEY: misc.feature
222 LOCATION: 16-190
223 OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 58

uuuuoacgaag ggcaacac ucgcenacac gcaguuuuc gugnnnnnn nagnccgaac au
ccnnnnnn ucccncac ccggacac ggggacac gcaguuuc uc
nnnnguucac uuuugcgccu gacuguguu gcg
nnnnnnnn gcccacgga ggcacuuggu aag

210 SEQ ID NO 59
211 LENGTH: 214
212 TYPE: RNA
213 ORGANISM: Vibrio vulnificus
220 FEATURE:
221 NAME/KEY: misc.feature
222 LOCATION: 16-190
223 OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 59

uauucaagcua gggcncacca uguuuuac cuuuuuncnnn ngg Nugcau a
nnnncacac uggugaaguac uuuugcgcac gcaacucguc ugcaacucu
aacncguc gcacacuc aacncugcc aacuncucu uuuuaccac
ugguuaaccu ucuuugggc uc
auguaaccu uuuuugggcu ggcacuuggu aag

210 SEQ ID NO 60
211 LENGTH: 23
212 TYPE: RNA
213 ORGANISM: Bacillus subtilis
220 FEATURE:
221 NAME/KEY: misc.feature
222 LOCATION: 12-131
223 OTHER INFORMATION: n = g, a, c or u

220 FEATURE:
221 NAME/KEY: misc.feature
222 LOCATION: 2, 13, 52, 53, 70, 92, 132
223 OTHER INFORMATION: n = g, a, c or u

220 FEATURE:
221 NAME/KEY: misc.feature
222 LOCATION: 2, 13, 52, 53, 70, 92, 132
223 OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 60

rwagagggc rannnnnnnna gua

210 SEQ ID NO 61
211 LENGTH: 237
212 TYPE: RNA
213 ORGANISM: Bacillus subtilis

<400> SEQUENCE: 61

auuucaagug uuaauggug aagauaggg ugcaacucuc aagguauugc
auuuggaaag aggguguguc uggagaugg gcaggguggc gagcgaaua
aaccccauc guuuuauu ucccuccggug cauggaauu auggauggcu gucgaauu

23
cauuuucug gaggcauauc uuguugcua uaauauuuua uguauuaa uugauaa

<210> SEQ ID NO 62
<211> LENGTH: 239
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 11
  <223> OTHER INFORMATION: r = a or g
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 78, 117, 177, 210, 232
  <223> OTHER INFORMATION: n = g or c
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 10
  <223> OTHER INFORMATION: v = g, c or a
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 123, 176, 211, 231
  <223> OTHER INFORMATION: w = a or u
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 167
  <223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 62

gaagaauagav rucgcagscuu caagauuaug ccuuuggaga aagauuauuu cuuguaaaa 60
ggcugaaggg ggcgcgugsc cgaagcaau aaaccctcaau cgguuauuu ugcuggguu 120
gowuugaua auuguauaggg uguuauaaa ucauauucuu ggagggyaaau cuuguuauuc 180
auaauauuu auagauuuau augauaaggg auugagagauu uuccucucuu wcuuuuuu 239

<210> SEQ ID NO 63
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 31-68
  <223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 63

cauuccuuuc guauauacuu ggagauaggg nuccaggaguu uccuaccaga uaccguuuaa 60
ugauucnauc uaugagggug ga 82

<210> SEQ ID NO 64
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 31-68
  <223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 64

cauuauuuuc guauauuggc aggaauaggg nucccagaguu uuccuaccaag cuacguuuaa 60
uagcuugcnuc uccagaaaaa au 82

<210> SEQ ID NO 65
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 65

aaaguaccuc ausaaauccu gggsaauugg ncccaaaagc uucuaccuggc uaggcguuas 60
ucgccccgac uacgccccgaa ga 82

<210> SEQ ID NO: 66
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-67
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 66

aaccucacuc guuuuuuccu cuuuccaugg ngaugagggu cuuuccaggu ucccccaaaas 60
uacuunncac uacgccccgaa au 82

<210> SEQ ID NO: 67
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 67

aaaacacuc guuuuuuccc gggsaauagc ucccccaaaagc uucuaccagg ggccccgguuas 60
cacugccgac uacgagggaa ac 82

<210> SEQ ID NO: 68
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 68

aguguaaacc guuuuuuccc gggsaauagc ucccccaaaagc cuuuccagagg cuuccggguuas 60
uacuccnegac uacgguauoc uu 82

<210> SEQ ID NO: 69
<211> LENGTH: 62
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 69

acacgaccuc auuuuuuccc gggsaauagc ucccccaaaagc uucuacccgg ccuccggguuas 60
uacggccgac uacgccccgaa ag 82
aggaacacuc aaauaauagc guggauaugg uccagcccaag uuuacccggg cacccgaa

nugucgcnac uaugggugac ca

agacaauuccu guaaacaugc aguaauaugg uccuauuugg uucuuaccugg uuacccguaa

aaacuacgac uacacagaaag uu

auuuaccacuc uuauuuaccuc aaauaauaugg uuucgggugg gcuuaccagg aancocgaa

auccuuggnau uccaaauuuu gu

uaaaauuuucu guaaancacc gguauaugg uuccgagaag uuuuaccuugc uguccauaaa

nugucgcnac uauggggggguu uac

caaaauacc guaaauagcuu agaauaugg uccuaaqgcuu cuuacccggu cuucgccgua

---continued---
<210> SEQ ID NO 75
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 75
aguuuaauc cuaaaauuuc cugaauugg uuccaggaugu uucuacaagg aaccuu aaa
60
nuuc cuuac ac uugagagau uu
82

<210> SEQ ID NO 76
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 76
usaauggauc guaauaugcgc gacgauugg ugguggau uucuacuagg aagccugu
60
cauucuanac uacgaaaua
82

<210> SEQ ID NO 77
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 77
usaauggauc guaauauguc gugaauugg uuccaaaggu uucuaccu uacccu aaaa
60
auacgacnc uacgagagugu
82

<210> SEQ ID NO 78
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 78
aaacaaaccuc guauaauaagg guugaagggu uucuaccagg aaccuu aaa
60
nuucocgnuc uauagugaa
82

<210> SEQ ID NO 79
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 79
aaacaaaccuc guauaauaagg guugaagggu uucuaccagg aaccuu aaa
60
nuucocgnuc uauagugaa
82
<400> SEQUENCE: 79
auuuugcuuc guauacucu auauaugg nauuagaggu cuuaccaag aancgagaa  60
nuacugnau uacgaaaaa gc  82

<210> SEQ ID NO 80
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Fusobacterium nucleatum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-61
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 80
auuuauacuc guauanagcc uauauugg nnaaggguu ucucuagguu aancauaaa  60
nuaacaca gc uacgaaaaa gu  82

<210> SEQ ID NO 81
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Lactococcus lactis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 81
acaaacauau uauauannccc uaggaugg nnnugggcgu ucuaccucg uanocguaaa  60
nucgcagnac aausagggaa uu  82

<210> SEQ ID NO 82
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 82
uaauauaguc guauaaguu ggauuugg naccggucgu ucuaccagg caaccguaaa  60
augacgaguc uacgacguuu ug  82

<210> SEQ ID NO 83
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 27-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 83
cgaaauccu guauauauug ugcgauuugg nucggcaggu ucuaccugg uucugguu  60
uacgcnacu uauagaauguu uu  82

<210> SEQ ID NO 84
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 94
aaugccuuc guauucuc guauauugg nuucgaasgu uucaacgugg uaccguuaas 60
ugaucugac uauggaagca ga 82

<210> SEQ ID NO 95
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 95
auagaaugc guauauuua ggggauuugg nuccacaggu uucaacacaga ccaccguuaas 60
ugguuugac uaugcagguua uu 82

<210> SEQ ID NO 96
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 96
auagauaccuc auauauuuu gagaauuugg ncuuacaggu uucaacccag caccguuuaas 60
ugccuguac uaugagggaa ga 82

<210> SEQ ID NO 97
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-66
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 97
uauguuuuuuu uauuaauucgc ggggaaugg uuucacaggu uucaacgguu uaccguuuaas 60
ugaaocgnac uauggaaaaa cg 82

<210> SEQ ID NO 98
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 98
acaauaaccuc aauauaucua aagaaauugg cuuugaaggu uucaaccaug uugoccuugaa 60
cgcacagnc uaugguuaac aa 82
<210> SEQ ID NO 99
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 68
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 99
uuauagacuc auuauaucua gagaauugg cuuuagaaggu uucaacgugg uggccauaag

<210> SEQ ID NO 90
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Streptococcus agalactiae
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 16-67
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 90
ugauuauacu auuuauagcgu gagauugg nnccuacgcu cuuccacaaga canccgunaas

<210> SEQ ID NO 91
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Streptococcus pyogenes
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 16-67
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 91
ugacacauccu auuuauagcgu gugaauugg nnccagcagcu cuuccacaaga canccnuuaas

<210> SEQ ID NO 92
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Streptococcus pneumoniae
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 16-67
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 92
cguuuauacuc guuuauaguc gugaauugg nncaacagcgu uucaacgagg ugnccnggaa

<210> SEQ ID NO 93
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 31-66
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 93
gagaacacuc auuuauuucg gagaauugg ncucgggagcu cuuccacggaa caacgguaaas

<210> SEQ ID NO 94
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Other
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 51-82
  <223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 94
ucuauuacuc gcuaauugg gagaauugg nnccagcagcu cuuccaqag uggccauaag
uugucgnauc uaugagugaa ag

<210> SEQ ID NO 94
<211> LENGTH: 82
<212> TYPE: RNA
<213> ORGANISM: Vibrio vulnificus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 31-68
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 94
ucasgcuuc auauauuccu uaugauugg nuuugggagu uucuaccaag agnccuuaaa
noucuagnau uaugagucu gu

<210> SEQ ID NO 95
<211> LENGTH: 69
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-69
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 95
nnumuauan nnnnnnnrrau augynnnnn ngunucuace nnnnnnccgu saannnnnng
acuaygrnn

<210> SEQ ID NO 96
<211> LENGTH: 201
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 96
gggauauua uaggaacacu cauauaucg cgugauaug gcagcaagu uucuacccgg
cacecuauau guggcuacau uggagcaeeu agugugccga cugtuacggc uuuuugugau
aucaaguauug cuuugcuuu uuuugacggc gcuaugcuuu uuuaauuccu uuacggagg
uagacaggau ggaucacauu a

<210> SEQ ID NO 97
<211> LENGTH: 93
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 20
<223> OTHER INFORMATION: k = g or u

<400> SEQUENCE: 97
gggauauua uaggaacacu cauauaucg cgugauaug gcagcaagu uucuacccgg
cacecuauau guggcuacau uggagcaeeu agugugccga cugtuacggc uuuuugugau
aucaaguauug cuuugcuuu uuuugacggc gcuaugcuuu uuuaauuccu uuacggagg
uagacaggau ggaucacauu a

<210> SEQ ID NO 98
<211> LENGTH: 84
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 19, 32, 44, 58, 59, 73, 74, 82, 83
<223> OTHER INFORMATION: s = g or c

<400> SEQUENCE: 98
gggauauua uaggaacacu cauauaucg cgugauaug gcagcaagu uucuacccgg
cacecuauau guggcuacau uggagcaeeu agugugccga cugtuacggc uuuuugugau
aucaaguauug cuuugcuuu uuuugacggc gcuaugcuuu uuuaauuccu uuacggagg
uagacaggau ggaucacauu a

<210> SEQ ID NO 99
<211> LENGTH: 83
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 18, 25, 26, 33, 43, 84
<223> OTHER INFORMATION: w = a or u

<400> SEQUENCE: 99
gggauauua uaggaacacu cauauaucg cgugauaug gcagcaagu uucuacccgg
cacecuauau guggcuacau uggagcaeeu agugugccga cugtuacggc uuuuugugau
aucaaguauug cuuugcuuu uuuugacggc gcuaugcuuu uuuaauuccu uuacggagg
uagacaggau ggaucacauu a
<400> SEQUENCE: 97

gguaaauoa uagagaaacw ke cauaawuac cau w uauw gcwgaaatug gcwegcaag u uuuacccugg 60
cacgguaua guagacguau gswgagcaaa ugg 93

<210> SEQ ID NO 98
<211> LENGTH: 51
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 8, 13, 14, 26, 32, 33, 37, 41, 42, 50, 51, 54, 55, 63, 67
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 18, 30, 44, 53, 60, 71, 72, 78, 79, 84, 87
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 17, 25, 34, 60, 74, 75
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 98

ycuaacucag agmnggyr g ggagyngcc cmnynanrec nncnrgcact n 51

<210> SEQ ID NO 99
<211> LENGTH: 251
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 150-251
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 99

ggacuuccgc acaacacaaau ucuacuacgc uuccuaccaau gagagcgacag gggagcuccgc 60
cgagacgaug ucaacacacgc gguuacuccg guacagccag gaucaacagug cuuaacacgc 120
cagacugccg cacugggacg guaagaagca gnmmmmmmmm mmnnnnnnn mmnnnnnnnmm 180
nnnnnnnn mmnnnnnn nnnnnnnnn mmnnnnnnnn mmnnnnnnnn mmnnnnnnnn 240
nnnnnnnnn n 251

<210> SEQ ID NO 100
<211> LENGTH: 124
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 106
<223> OTHER INFORMATION: k = g or u

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 13, 14, 46, 47
<223> OTHER INFORMATION: r = a or g

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 13, 14, 46, 47
<223> OTHER INFORMATION: r = a or g

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 19, 42, 97
<223> OTHER INFORMATION: n = g or c
<220>  FEATURE:
<221>  NAME/KEY: misc_feature
<222>  LOCATION: 94, 85
<223>  OTHER INFORMATION: y = c or u

<400>  SEQUENCE: 100

ggguccuwwa carragwsc agaggacug gcgcacgaa gswvcrcaas cguguaau 60

aguguacagc caugaccaag gugyyaauuc caagaevuc gaacakcuug gaagawaga 120

agag 124

<210>  SEQ ID NO 101
<211>  LENGTH: 245
<212>  TYPE: RNA
<213>  ORGANISM: Bacillus subtilis

<220>  FEATURE:
<221>  NAME/KEY: misc_feature
<222>  LOCATION: 186-245
<223>  OTHER INFORMATION: n = g, a, c or u

<220>  FEATURE:
<221>  NAME/KEY: misc_feature
<222>  LOCATION: 149, 160, 177
<223>  OTHER INFORMATION: n = g or c

<220>  FEATURE:
<221>  NAME/KEY: misc_feature
<222>  LOCATION: 148, 161, 176
<223>  OTHER INFORMATION: v = a or u

<400>  SEQUENCE: 101

ggucagaaaa auggaaactc uaaauucuaa ucgugagagg uggaggagcu gcucuuaa 60

aaccuacca aocgguuug uugcaauug caaagcgca auggcuuaaa uccagaacgc 120

guuuuuuuu ccuugagac acaagagwsc guuaccccsc uuuucuuuaa gaagawgggg 180

uuuunnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240

nnnnnn 245

<210>  SEQ ID NO 102
<211>  LENGTH: 167
<212>  TYPE: RNA
<213>  ORGANISM: Bacillus subtilis

<400>  SEQUENCE: 102

gguacauuc aaaaaucuaa caaagcgcc uggaggacuc gaccuaugaa gcocggcaac 60
cugcaauug uguagaguc uacuucacgc aaaauauuu ccaauuugga aguauagggc 120

ugucaugcg uccugcuuu cuuuccgcgc gauggasagu uuuuuuu 167

<210>  SEQ ID NO 103
<211>  LENGTH: 160
<212>  TYPE: RNA
<213>  ORGANISM: Bacillus anthracis

<400>  SEQUENCE: 103

ggagcguuaucc aagagatagc gaggcaguc gcocggcga gcocggcgaac cugcuuaag 60

aaagaacgg gcguuuccaa gcacsugga accuuuugg aagauaugg uaaaauuuu 120

uaccgcaga uuuuuuuca auggggaga uuuuuuuuuu 160

<210>  SEQ ID NO 104
<211>  LENGTH: 80
<212>  TYPE: RNA
<213>  ORGANISM: Bacillus subtilis
acacgacccu auauaauccu gggaauugg cccauaaggu uuuauacccgc aaccoguaau 
60
ugccgacua ugcggagaag 
80

<210> SEQ ID NO 105
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 52-60
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 105
aggaacacuc auauaauccg guggauugg caagcaaggu uuuauccggc aaccoguaa 
60
ugccgacua ugggagaac 
80

<210> SEQ ID NO 106
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 52, 60
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 106
auauacucu auauaaccuc auauaauugg uuugagggug uuuacccagga aaccoguaa 
60
auucugauua caasauuugu 
80

<210> SEQ ID NO 107
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 52, 60
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 107
auuugcucu guauaaccuc auauaauugg auuagagguc uuaaccaga aaccogaga 
60
auucugauua gcagaagaa 
80

<210> SEQ ID NO 108
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Vibrio vulnificus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 52, 60
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 108
ucacgcuuc auauaaccuc auauaauugg uuuggaggug uuuacccaga gaccuua 
60
ucuucugauua auuagc 
80

<210> SEQ ID NO 109
<211> LENGTH: 80
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<400> SEQUENCE: 109

cacucaauaa aucgcgugsa uuuggcacgc aeguwwcuac cgycacagyu aaaugucga
60
cuauggug
69

<210> SEQ ID NO 110
<211> LENGTH: 63
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 110

uuguaauacc ucauaauauu gguuuaggg ugucuaccag gascoguaas auccuaua
60
caa
63

<210> SEQ ID NO 111
<211> LENGTH: 102
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 111

uuguaauacc ucauaauauu gguuuaggg ugucuaccag gascoguaas auccuaua
60
caaaauuugu uuauggacauu uuuguaauac aggauuuuu uu
102

<210> SEQ ID NO 112
<211> LENGTH: 466
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE: misc_feature
<221> LOCATION: 21-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 112

atatcgttc ttatacaag nnmaagcaga gggannctgg nnmmncgac gcagottmm
60
agcaacgg tgaatggnnn nnmnmnmnm nmngntcann nnnnnnnnn
120
nnnmmmmnn nnmmngcatt gaccaagctg ctaattcaca gmmmmnnmnmnmnm
180
nnmmncgaca nmnhmmmmmn ngctttcgaag atasgaagag acaaattcac tgacaaann
240
nnnnmmmmnm nnnnnmmngt ottttnnmnm nnmmnmmnmnmnmnmnnnmnnmnmnnmnn
300
nnnnmmmmnm aggaattttt tatttcttt tttttctttg tgtggtgat aagggagcag
360
gacattgggs cttttgaag attgcaaaag acagttgtta atcggtgac gcgcctagg
420
gacgctocct tactctatag gcattgacag gtgttctggag gacgctcaata tttcaagcc
480
gggag
486

<210> SEQ ID NO 113
<211> LENGTH: 466
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE: misc_feature
<221> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 113
tccatatcctc ttatactttag gnnnggtgga ggganncttg gnnmccctata gaaacctnmc
60
agcaacggc tgggtttgtcnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnn
120
nnnnnnnnn nnnnngaag cggccaaagt gtaaatncca gmnmmnnncna gogtnnnnnn
nnnntttttt nnnnnnnnnt tgttttgaag ataaagagaag cggtnaaann nnnnnnnnn
nnnnnnnn nnnnnnnnn nnnnntcccc cttttttcnn nnnnnnnnt tatnmmnnnn
nnnnnngaga agggggttttt atttttaaaa ggggaaggtgt cagctatatg ttcaacaacg
ttgaaacgaa attagotcga attgggaacc gtagcgtga aagtcacggga acagtggaatg
cctttatcca tttaaacaac gcataccgcc acagagggat cagagaattc accgagattg
attatg

<210> SEQ ID NO 114
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: mloc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 114
acatcttttc ctatcagag nnntgggcca gggattttgg nnnnnctttt gacccaaanc
agcacaagcc cmmnnnntga ataacattgtg gaatgggagg gcaagcttttt tggccgag
acgtgtctct cataanmmnn nggcaacgggt ctaattttcc a tmnnnnnncag atnnmmnnnn
nnnnntgtnnn nnnnnnnnnn ngtttcagag attagagggg cagttttttta cgtagaaan
nnnnnnnnnn nnnnnnnnn nnnnnnngc ctttttctcn nnnnnnnnt catnnnnnnn
nnnpnngggaa gaggcttttt gttttggaaac acctcttagt cagccctgtat cggccggtga
aagagagttc tttaaatata aagaggagaaa aacaaggaaca ccatcacaat catagaaattt
aggatttcgg aagaatgacc tgaaccgggga atggaaaaac gcagctgag cgtatttgga
agcgc

<210> SEQ ID NO 115
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: mloc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 115
atatatcttc ataatcagag nnnntgggcca gggattttgg nnnnnctttt gacccaaana
agcacaagcc cmmnnnntga ataacattgtg gaatgggagg gcaagcttttt tggccgag
acgtgtctct cataanmmnn nggcaacgggt ctaattttcc a tmnnnnnncag atnnmmnnnn
nnnnntgtnnn nnnnnnnnnn ngtttcagag attagagggg cagttttttta cgtagaaan
nnnnnnnnnn nnnnnnnnn nnnnnnngc ctttttctcn nnnnnnnnt catnnnnnnn
nnnpnngggaa gaggcttttt gttttggaaac acctcttagt cagccctgtat cggccggtga
aagagagttc tttaaatata aagaggagaaa aacaaggaaca ccatcacaat catagaaattt
aggatttcgg aagaatgacc tgaaccgggga atggaaaaac gcagctgag cgtatttgga
agcgc
<210> SEQ ID NO: 116
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 116

catatatttgc tatacaagag canggagcag nnncccgag gaagccnnnc 60
ggcaacgcac ttnnnnmmnn nnnnnnnnn nnnnnnnnn nnnnatannn nnnnnnnnnn 120
nnnnnnnn nnnnnnnnn aagcaaggtg ctatncttt gnnnnnnnccag ctnnnnnnnnn 180
nnnnnnnagnn nnnnnnmmnn nggcgtcagag ataaagctcg gcagcgaacctc gaannnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnc otcttagagc cnnnnnnnnnn attnnnnnnnnn 300
ngcagttgta agaggttttt tgtattgtgat gaaatgsaa ggcgtccttg catgagtcgag 360
ttstatcgc cattctcctt cagcgaaccgg gagcccgata cagagaaggagacgaacaa 420
atgcagcagc gattgacagt aggcgcctgg attgacatgc cctctgttaa acagggccaa 480
atgcca 486

<210> SEQ ID NO: 117
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 117

atctaaaaac tgtacagag cnnmgtcga ggganctgg anmocctnat gaagccnnnc 60
ggcacagctg anmmnnnnn nnnnnnnnn nnnnnnnnn nnnntagttnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nngtgaaggtg ctnacttcca gnnnnnnncaa aatgnnnnnn 180
nnnmtattn nnnnnnnnnn attttgaag ataaagctgct cttgctgttc tctgnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnntcct cttctctctnn nnnnnnnnnnn gcctnnnnnn 300
nnnnmmgtt gaaatggtttt tatattcga ggttaaaaggg cttatctgtat atcagcgcoc 360
ggacatacc aaagcagaca gcgacaaa gaagctggtg cttctctcgg 420
aacggagaa ccaagataag ttaaaattgta taaattggaa gaacccgca acaaatcttcc 480
ggaga 486

<210> SEQ ID NO: 118
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 118

tcaatatctt ctatccagag nnnaggtgga ggganntgg nnnncctatat gaaacctnnnc 60
ggcaacanmn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnntatnnnn nnnnnnnnnn 120
nnnnnnmn nnnnnnnnn nnnnnntgtg ctaatttnca gnnnnncnaa gnnnnnnnnn 180
<210> SEQ ID NO 119
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-307
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 119

gcgtgatgcc ttacccctgg gtgggggcaag gggggcggcc tggggcctgt ggggggctgg ggggggccgc gggggcggcc 60
ggggggctgg ggaggggggg ggggggctgg ggggggctgg 120

<210> SEQ ID NO 120
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-305
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 120
ttcttttttg gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc 60
gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc gggggggccc 120

<210> SEQ ID NO 121
<211> LENGTH: 486
-continued

<212> TYPE: DNA  
<213> ORGANISM: Bacillus subtilis  
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: 22-305  
<223> OTHER INFORMATION: n = g, a, c or t/u  

<400> SEQUENCE: 121  

aagttgtaco ttatacagag anmnagtcgg gggannctgg nnncoctnat gatacnnmc 60  
gggacocgt gttmnnmmnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120  
nnnnnnnn nnnnnnnnn cagaatgtgtg ctaatntcct tmnnnnnaag aacnnnnnnn 180  
nnnnattgcnn nnnnnnnnn gtttctgcag atagyccgaa gatttctgct ttcaannnnn 240  
nnnnnnnnn nnnnnnnnn nnnnnnnngc ttctcttnnn nnnnnnnnm caacnnnnnnn 300  
nnnnmaagga agagcttttctt atacctgtaa tatattgcagaa aagagcggaa taacatggt 360  
caacaaacga atgtgctcag aacaaaaaca aaacaaacac gcaagaccac ctcgcagccgcccc 420  
gacaattgtcag acggcgtcga tgcgtgaaac gcggccgcac aaaaaagctgcct 480  
gctgtg 486  

<210> SEQ ID NO 122  
<211> LENGTH: 486  
<212> TYPE: DNA  
<213> ORGANISM: Bacillus subtilis  
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: 22-305  
<223> OTHER INFORMATION: n = g, a, c or t/u  

<400> SEQUENCE: 122  

aagtttttcc ttatacagag anmnagtcgg gggannctgg nnncoctgcg gatacnnmc 60  
gggacocgt gttmnnmmnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120  
nnnnnnnn nnnnnnnnn cagaatgtgtg ctaatntcct tmnnnnnaag aacnnnnnnn 180  
nnnnntgann nnnnnnnntt gctttgcag ataggttgca gtttgcagc caannnnnnnn 240  
nnnnnnnnn nnnnnnnnn nnnnnnnngc ttctcttnnn nnnnnnnnm caacnnnnnnn 300  
nnnnmaagga agagcttttctt atacctgtaa tatattgcagaa aagagcggaa taacatggt 360  
caacaaacga atgtgctcag aacaaaaaca aaacaaacac gcaagaccac ctcgcagccgcccc 420  
gacaattgtcag acggcgtcga tgcgtgaaac gcggccgcac aaaaaagctgcct 480  
ggcggaa 486  

<210> SEQ ID NO 123  
<211> LENGTH: 486  
<212> TYPE: DNA  
<213> ORGANISM: Bacillus halodurans  
<220> FEATURE:  
<221> NAME/KEY: misc_feature  
<222> LOCATION: 22-306  
<223> OTHER INFORMATION: n = g, a, c or t/u  

<400> SEQUENCE: 123  

tatatatttt ttatacagag ttntmggtcgg gggannctgg nnncoctgtg gagaconnmc 60  
gggacocctt tmnnnnmmnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120  
nnnnnnnnn nnnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnnn nnnnnnnnnn 180  
nnnnntgann nnnnnnnntt gttttgcag atagaagtgcg ccaatnccc gnnnnnnnccag aacnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnncac gtccttcnnn nnnnnnnnt tatcnnnnnn
300
nnnnnnngaag aggtggtttt tcggggattta acacottaac tgcgggaag attacgtgtt
360
atgtacccag aacacgaaaa caaaaaaag aacaacttgg aatgaggagg ggttgtcatc
420
gaasaaattta tcgcttatttc acaagaaaca tgtaagagag gttccactta ttaacgtact
480
tgagg
486

<210> SEQ ID NO 124
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 124
ataaaaaacgc ttatcgagag anngggcaga ggganctga nnncccgag aaagccmnnc
60
ggcacccgt ttcgtnmnnc nnnnnnnnn nnnnnnnnn nnnnnncnnn nnnnnnnnnnn
120
nnnnnnnn nagcaacgac aggtgccat tntcagnnnn nncagaattgn nnnnnnnnnn
180
tttnnnnnn nnnccattct gaaagtaatgg cgagggcag aannnnnnnn nnnnnnnnnnnnn
240
nnnnnnnnn nnnnnnnnn nnnnnnncc ttctccnnnn nnnnnnnnt tatcnnnnnn
300
nnnnnnnnng aaaggttttt tttgtaagaga gccaagtttt tataaaatag tggaggtggcc
360
atagaaaggg gaaataaatcg agtgggtgtg atggatattg gcacggttg
420
gcgagtgttt tgcgagagtc tagtcttaatt ggacgcggaga ttaaggaaag aagttactct
480
cgaat
486

<210> SEQ ID NO 125
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-302
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 125
ttcgctattc ttatcgagag nnnaggtgga ggganacgg gnnccgga gaaacctnmnc
60
agcacaacgc cagcagnnnnn nnnnnnnnn nnnnnnnnn nntecmnn nnnnnnnnn
120
nnnnnnnnnn nnntntntntgt gntcaggtg ctaatncct gnnnnnnncag gcanmnnnnnn
180
nnntntattn nnnnnnnnn ngtctggag atagaagagaa gcagtgsgaa tcacnnnnnn
240
nnnnnnnnnn nnnnnnnnncac cccttcctt ctntnnaaacct tatacgacnnn
300
nnngagaaggt aggtcttttt ttacacaact agaaagata gacttcttcag ttagtttaa
360
gaaaaatgaa ggttggcag acttggccac gagctgtatt tttcataagaga gtagtatgg
420
gacacaccc gaacctgaa gaaagaaac cagttttagc acaatcgaga aatgaaatgg
480
atgac
486

<210> SEQ ID NO 126
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 126

acgatatacc ttatccagag tttmgttga gggannccag ngnnncggaa gaaaccnncc
60
agcgaacac scotnnnnnn nnnnnnninn nnnnnnnnn gnttasacca nnnnnnnnn
120
nnnnnnnn nnnnnnnngn ygaaxaggt ctanmncct gnnnnnnnca ggcnnnnn
180
nnnnngtnm nnnnnnnnn gctttgaag ataaagagg aagagtagtg taatatttmm
240
nnnnnnngaa aaggtttttc ctactatttta tactttttca agttgctgtg gggaaggtgg
300
tgctgatatca tgttattggc agctctgctg tgttaaggt gttgattaagg gaggatatcc
360
gtaatggcga gtataagaa gttgtagttc atttatattc aagttggttag cagaaggaca
420
tcoctga
486

<210> SEQ ID NO 127
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 127

agcagaacac tttagttcgag ggannntggc gnnncggat gaagcnnnc
60
agcgaacagc aagcnnnnnn nnnnnnnnn nnnnnnnnn nnnngaatcn nnnnnnnnn
120
nnnnnnnn nnnnnncct gaaaggttg ctattnccct gnnnnnnnca agcnnnnnn
180
nnnnngtatnn nnnnnnnnn gttttgagag atggagaag gggaagcgtg aaacattttm
240
nnnnnnnn nnnnnnnnn nnnnnnnct tctctgcnnn nnnnnnnct catgmnmmn
300
nnnnncgcgg aagagtttttt tgttattttt attcagttttg attacgagaa ttgtaattttc
360
ttacgatat gattgtttgg cttccttcgaa cgaattaagc agagttttgg gttttttgc
420
tttggtatatca ttgtttggc gttgagagaac gttatatttt ttagaagaa
480
gggtt
486

<210> SEQ ID NO 128
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 128

atagttgagc ttatccagag nnnagatggc gggannntgg cgcggcgcat gaaagctnnnc
60
agcagcagc ottnnnnnnn nnnnnnnnn nnnnnnnnn nnnngatatnn nnnnnnnnn
120
nnnnnnnn nnnnnnnnn aagttgtagtg ctattncca nnnnnntag gttinnnnnn
180
nnnttacann nnnnnnnnn nnnncgctattaag atacaaaaag ctaatagttt taaaannnnn
240
nnnnnnnn nnnnnnnnn nnnnnnnnn cttctttcnn nnnnnnnnta cttctttmnn
300
**SEQ ID NO 129**

**LENGTH:** 486

**TYPE:** DNA

**ORGANISM:** Oceanobacillus iheyensis

**FEATURE:**

**NAME/KEY:** mloc_feature

**LOCATION:** 21-306

**OTHER INFORMATION:** n = g, a, c or t/u

**SEQUENCE:** 129

```
atgacaaactctattacagagntggaagggacctgggntggtccagggaaagctg
60
ggcaaacagctcactattaaatctcctgggagggagtctcgactgattttgtgtctagcc
120
ctctgcattcattcaccagctgagctgtgtagagctgtgtgactgagtgagggagtctc
180
ctctgcattcattcaccagctgagctgtgtagagctgtgtgactgagtgagggagtctc
240
ctctgcattcattcaccagctgagctgtgtagagctgtgtgactgagtgagggagtctc
300
ctctgcattcattcaccagctgagctgtgtagagctgtgtgactgagtgagggagtctc
360
tctactgagctgagctgtgtagagctgtgtgactgagtgagggagtctc
420
tctactgagctgagctgtgtagagctgtgtgactgagtgagggagtctc
480
tgcgac
486
```
<222> LOCATION: 21-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 131

ATGAAATACTTATCCAGAGNMTGACGAAGGAANCAGAMNNCOCTANCGATGTCANC60
AGCAACCTCCTMNNNNNNNNNNNNNNNNNNNNNNNNNNTTCCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
-continued-

ggagagttaag agggaggaata aagattgata tcctatcgaag ggttaagctaa agtatatttca 420

atttat 486

<210> SEQ ID NO 134
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 134

tcttgaaactctactagag aaaaaggggga gggagaataag nnnncccaac gaaaccctnmc 60
ggcacaggt tcnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
acacaaagcgaattgcgtagctaactggtcga attggaattct acacaaagcgaattgcgtagctaactggtcga 480

<210> SEQ ID NO 135
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 135

atgaatatc tctatccttag aaaaaggtgga gggagaatgg nnnnccccaaa gaaaccctnmc 60
ggcacaggt tcnnnnnnnn nnnnnnnnn nnnnnnnnn nntagctnnn nnnnnnnnn 120
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnncnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
tgctggcatt ttgaggggtg gtgggtgtatgt gtagatattgctgcag tctaatagc 420
tttctactttt gggagatatt tcatagagcag gaaaccctnmc 486

<210> SEQ ID NO 136
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-308
<223> OTHER INFORMATION: n = g, a, c or t/u
-continued -

<400> SEQUENCE: 136

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Location</th>
<th>Other Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>ttatatatcc tttccagag tnnnngggga gaatntcgg nnnntccatt gatccgccc</td>
<td>60</td>
<td>22-306</td>
<td>n = g, a, c or t/u</td>
</tr>
<tr>
<td>agcgaacctgt taannnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnngtagnn nnnnnnnnn nngcctgatg atgagaaag tgggttcac tttttaattaa</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaaaaa</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnnnngl aagcgatctga tctttgtgct gagntgcggt nnggagaaat nnggggggt cggacaagtc</td>
<td>360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ggtggacttc aaatataggt aacagatctga gttctgaag atggggggt cggacaagtc</td>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaggtgatgct agtttatataa aatctatataa gaaaggggaga ggggtgttata taaggcattag</td>
<td>480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaaaacc</td>
<td>486</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 137
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 119-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 139

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Location</th>
<th>Other Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>atgaaatcct ttatctgag nnnaggtgga gaagaaaag tnnnccaaa gaagctccmc</td>
<td>60</td>
<td>21-304</td>
<td>n = g, a, c or t/u</td>
</tr>
<tr>
<td>gggcaactgt tcnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnnnnn nnnnnnnnn gaaaactgtc ccnaatncna tnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnntctnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnaatgag nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnntgattg</td>
<td>360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttctgattt ccgaacctga aacagatctga ggctgatcct gcttgccgatc cggacaagtc</td>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gacaggaactgc aacagatctga gggctgctgct cggacaagtc aacagatctga ggctgatcct</td>
<td>480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>atggtg</td>
<td>486</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 138
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 139

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>Location</th>
<th>Other Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>atgaaatcct ttatctgag nnnaggtgga gaagaaaag tnnnccaaa gaagctccmc</td>
<td>60</td>
<td>22-306</td>
<td>n = g, a, c or t/u</td>
</tr>
<tr>
<td>gggcaactgt tcnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn</td>
<td>180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnngtagnn nnnnnnnnn nngcctgatg atgagaaag tgggttcac tttttaattaa</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaaaaa</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nnnnnnnngl aagcgatctga tctttgtgct gagntgcggt nnggagaaat nnggggggt cggacaagtc</td>
<td>360</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ggtggacttc aaatataggt aacagatctga gttctgaag atggggggt cggacaagtc</td>
<td>420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaggtgatgct agtttatataa aatctatataa gaaaggggaga ggggtgttata taaggcattag</td>
<td>480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aaaaacc</td>
<td>486</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
tcctactttt gaaattgcaa aacaggttaat tcagcacgcc gaagaatggg gataatgatac 480
gaccttt 486

<210> SEQ ID NO 139
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-300
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 139

ttataacattt tctatcagag nnnaagctaa gggacnctgg nnnnccccgt ttacagttmnc 60
agcaccctct amnnnnnnnn nnmnmmmnmm nnnnnnnnnt nnnnnnnnntnnnnnnnn 120
nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
ctaattatct ttaattgagag tnnngggtgaa gggannntaggg nnnnccctgt gaaagcmnnnc
60
ggcaacggtct cactgtnmn mnmmnmnmnm mnmmnnmnmm mmmmtctcmn mmnnmmnnn
120
mmmmmmmm mmnnngtta tgaattaccg ctaattnctc gmmnnnncnca actacnmnmmn
180
mmmmgcacan mmnnnnnngtm atttggagaa ataaagaggg tttgattacgtt cctacgctg
240
gtnttgtaat mmnnnmmnmn ctnctctgatnm mmnnnmmnt tcaonmmmmnm
300
mmnnntagggg aggcatctttt tgaattcccccg gaaatattca caacatgaga aaagaggaag
360
gattatctc carctgctac gtaaaagggcg ctnccgggtc ctacagtctg gggggggtgag
420
tctggagaa aatccctgta ctgctgggag aacgtgctagt taactcctata caagttgtcg
480
cagggga
486

<210> SEQ ID NO 142
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-302
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 142

ataatttttc ttaataagag tnnnggaggg ggganntaggg nnnnccctat gaaagcmnmnc
60
ggcaacgctcc aannnnmmnn mmnnnnnnnn mmnnnnmmnm nnttattnnn nnmmnnnnnn
120
mmmmnnmm mnmmnnmmnm tgggagttct ctaattnctc annnnnnccag mmnnmmmmmm
180
mmnmnmtnn mmnnmmmmnm mmctgagag atgagaactt gtttttcccmm mmnnnmnnnnmm
240
mmmmmmmm mmnnnnmmnm mmnnnnnct cfncttttn tnnnnnnnnnt taattmnnnnnn
300
nnnggtattac agcattttttct tttttttnct attaaagggag gaggattaggcagagagag
360
attttgctaa ttttgagtaca aagacagtct cttggggaag agttttggtg agcagacagt
420
gggcaaatag ctaccggaat ttacacgtgac atccctttaat cctacgctgac tttaataaag
480
gaagctg
486

<210> SEQ ID NO 143
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 143
tgtaaaaacct ttaataagag tnnnggtgaa ggganntaggg nnnncccttt gaaaocnnnncc
60
ggcaacgctct atattnnnn mnmmnnmmmn mnmmnnnnmn nnnnttttnn mnmmnnnnnn
120
mmmmnnnnm nnnnnnaatt atatggctgt ctaattntctc gnnnnnncnag mmnnnnmmnm
180
mmmmncnn mmnnnnnnnn mmngttgagag atgagaacta taocggaasgta aannnnnnnn
240
mmmmnnnmnm mmnnnnnmgc cggagggmmn mmnnnnmmt attttmnnmmn
300
mmmmnncnca agggcctttttt attttctgact attttttttag ggggcatactt gtaatgactt
360
tcataataa attttgtaac taacacatct ttgaattctgctt attgcctgtg gggacgctgt
420
attcattacct ttaattttaga tgaaggggaac ttttaaggtt ctttatcttg tctacgctgt
480
tccaat

<210> SEQ ID NO 144
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 144
taatatttcc ttatacaagag nnnnaccgga gganncttg gg nnncaccaat gatggttmmn 60
gaccaacoag gtnnnnnnnn nnnnnnnnn nnnnntattn nn nnnnmmnnn 120
nnnnnnnnnn nnnnnnnnn acttatggttg ctaatttcca gnnnnnnccag gnnnmmmm 180
nnmnttattnn nnnnnnnnnn nntottgaag atgaggagcg actattttaaac cattttttattt 240
tttgttaatac mmnnnnnnnn nnnnnnnttc cttctctttm nnnnnmmnta tammnnnnnn 300
nnnnnagaaggg gaggattttta tttttgtaaat attagacacca cttctcttcattttttctt 360
attctattaa agtggttggt atagacacat ttttatattaa aagaagagaga aaattactcaca 420
attaccttc cttctacctc ataagctttat gaaattttccc caatctatac taaaaatatttt 480
ttacta

486

<210> SEQ ID NO 145
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 145
attatgtcag ttatacaagag annggttggga ggannccggg nnnncccttg gaagcnnnnm 60
gaccaactgtg atnnnnnnnn nnnnnnnnn nnnnntttattn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn atcaggttgg ctaattttct gnnnnnnccag nnnnnnnnn 180
nnnnncattn nnnnnnnnn nngotgagag atgagaaat aatgagatct tttttmmmmn 240
nnnnnnnnnn nnnnnnnnnng gggaggn nnnnnnnnttt tattnnnnnn 300
nnnnnctct cgtcttttat tttttttttta cttatggggaaaa aggtgtaatg acatttata 360
aatataaaat gtttttatattttttcagg ataattaggtt ctaataagatg tttttttctgg 420
gatataaggc ggagaataat ttgagattgt ttgctattaag ggagctggag aatcaactt 480
acttag

486

<210> SEQ ID NO 146
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 146
attatatttc ttatacaagaa nnnnggttggga gganncttg gg nnnccctat gaagccnnnt 60
gacaacgcggc nnnnnnnnn nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nngtacgcgtg ttaatnccct gnnnnnncaaa aaccnnnnnn 180
nnnttattn nnnnnnnnn gttttggag ataagaaaaac agottattaa ttaatgtaa 240
tgttaataan nnnnnnnnn nnnnnnnntc cgttccccnn nnnnnnnntt tatnnnnnnnn 300
nnnnnggaaat aagatttttt ttatatattt aaaaaattaa ttagaacggt gaaaaataatg 360
cctataaaaa tacctgataa tcttccagca gaaaaacctt taaatgaaga aatatattt 420
ttatgtgatg aggatagacg cttcatctca gataaagac cttttattat ttttatttgtttt 480
aacctt 486

<210> SEQ ID NO 147
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 22-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 147
tgataaggtc ttatcaagag annggtgga ggganctgg nnnccctat gaaaccnnnc 60
aacaacccg attttnnnn nnnnnnnn nnnnnnnn nnttttaattt nnnnnnnnn 120
nnnnnnnn nnnnnnagat atgtatgggtg ttaatnccct gnnnnnncaaa aaccnnnnnn 180
nnnttatnn nnnnnnnnn nttttgagag ataagaggat tataaatttt tagaagct 240
aannnnnnn nnnnnnnnn nnnnnnnntc cttcttccnn nnnnnnnnncta aannnnnnnn 300
nnnnnnngaa agggtatttaa ttttatatat ttttaggttt agatattttg gtttaaatat 360
aataaagg aggatttttaa aagattgtgaa gaaagaaaaattt gttttttgaa acacattcag 420
gttctacgc gacaagttgc tgcattcaact acaggtatca gaagcttcac ttatttataa 480
aacaca 486

<210> SEQ ID NO 148
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 22-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 148
atgaaactc ttatcaagag annggtgga ggganaggg nnncccttt gaaaccnnnc 60
ggcaacgcgt gttattttttt nnnnnnnnn nnnnnnnnn nnsatattnn nnnnnnnnn 120
nnnnnnnn nnnnnnagat cattatgggtg ccatatnccct gnnnnnnccag aannnnnnnn 180
nnnttattnn nnnnnnnnn nttttgcag ataagagaga gaatattaa nnnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnngttt cttcttccnn nnnnnnnnttattnnnnnnnn 300
nnnnnggaaag agggtacttta ttatatattg ttgagagaa gttttataaa ttagaaggt 360
aatatatct gatacagtaa cagagggcct cttgataaata atggctgac taaaatcaga 420
cggctattta gtagccatg tggaaaaaga tccaaagagga agagttgctt ggttaaactc 480
agtgac 486
ttataaactc ttcagcagc anngggctga gggaaasagc nnnncctat gaaacgmcnc
60
ggcacccagt gannnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
120
nnnnnnnnnn nnnnnnnmmnt cactaagcttg ccatttnnec gnnmmntaa agannnnnnnn
180
nnnnnmaatnn nnnnnnnnnnnt tcctttcaag atggagagaag ataaatttag tggtaacta
240
aaannnnnnnn nnnnnnnnnnnntc tccttttaaa tccttttttaa taannnnnnnn
300
agggtttgga agagacctttt ttataaccaaa aatatattta aaggggctgc taataaataa
360
gtttttaaat taagcccttta agatattttt ttgaaatcgtg gggaagataaa taagtttatt
420
ttgtaaatc aacggagttg gaaataaaat aatggaaggg ggtgaattag ctatctttatt
480
atgata
486

agcagcagc atttctttcgc gttttaaaca atgttgtatt taaaattaaa aatacagtag
120
taatcttct gtcgaaagggcttacagtct gtaattaacttct gnnnnnncgg tnnnnnnnnn
180
nnnnnnnnnn nnnnnnnnn nnnnnmct gataagaag gttcgtttaa gattaaaat
240
nnnnnnnnnn nnnnnnnnn nnnnnmct tctatcnmn nnnnnnnnn tcgtnnnnnnn
300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
360
ggagagaga aatgtgaaaa agggagttt tcacagttatt tacattatat aatggatttt
420
ttttttattg tggggatacct attagtaatg aatgattttct actctgatct ttttttttagt
480
ccagga
486

agcagcagc cmmnmmnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn
120
-continued

```
nnnnnnnnn nnnnnnnnn nngttctatg ctaattncg atnnnnnncag aannnnnnnn

  180
nnnnnnnn nnnnnnnnn atctcttgccg ataaagtaga gttttcataa gngnnnnnnnn

  240
nnnnnnnnnn nnnnnnnntgt ccctggattct gnnnnnncac aaaaaannnnnn

  300
nnnnnccaggg aagcgcgtatt ttttatttgcc ttaaagaggg gacgttttttgt tagatcaga

  360
aatttttatt agtgcgcgttt atctctggttt ttgctttggt gtaacggtgt tcgagcttggt

  420
tctgggctag ttcacacaaa gcaacagggc tggccaaagc gaaagacggt gcgtcttcca

  480
ttatg

  486
```

<210> SEQ ID NO 152
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-305
<223> OTHER INFORMATION: n = g, a, c or t/u

```
<400> SEQUENCE: 152

atatbtcttc ttactagag cnnngggaga ggannntcgg nnnnnccag taggcnnnnc

  60
nggcaacctaa ctttatnnnn nnnnnnnnn nnnnnnnnn nnttaacgnn nnnnnnnnnnn

  120
nnnnnnnnn nnnnnnnnta aagcgcgttt ataaatttcc gmnmmnncca aatggnnnnnnn

  180
nnntgttttn nnnnnnnnc gtttttggtg ataaagaggg ctcgtatgtg tctacatcccc

  240
nnnnnnnnnn nnnnnnnnn nnnnnnnncc ttctctattn nnnnnnncnc taannnnnnnn

  300
nnnnnaatag agaagttttt ttatgttttt tcatgaaataa tctgacataa cacacacaacat

  360
actagggag aaaaaagatg aaaaaatatt aaaaaaggggtt agagaaatatta tctgacataa

  420
gcgtcttttt agagattagc gcatgtgagg gaggcagttg cgataaaagcc ttaacacag

  480
aaaaaa

  486
```

<210> SEQ ID NO 153
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-303
<223> OTHER INFORMATION: n = g, a, c or t/u

```
<400> SEQUENCE: 153
tagattttt ttatcaagaa nnnngtgtaga ggannntcgg nnnncoccctt gaaagctnnt

  60
agcaacccga annnnnnnnn nnnnnnnnn nnnnnnnnn nntttatnn nnnnnnnnnnn

  120
nnnnnnnnnn nnnnnnnnnc ccctggattct gtaattncca gnnnnnnncag nnnnnnnnnn

  180
nnntatattn nnnnnnnnn nnnntgaaag ataaagctga aatccaggtt taggaaacct

  240
tatnntnnnn nnnnnnnnn nnnnnnncc tctctgcggg nnnnnnccct atatannnnnn

  300
nnnctgcag ggtgtttttt tgaatggaaat tactgttataa cacatatcga agagagaaggg
gttttggagat agaatgaagat ataatgtgtat aatctgaaaaa aatcagaggtt cacaacacgg

  360
tttatagtc ggcagtttatt gttatggaaat tactgttataa tacatatcga agagagaaggg
gttttggagat agaatgaagat ataatgtgtat aatctgaaaaa aatcagaggtt cacaacacgg

  420
tttggtttt tcagagtaac acattctaga ggcgtacatta tttatcaacac gacgtcatac acaattgata

  480
gcccgg

  486
```
<210> SEQ ID NO 154
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 154

acagatgta ctatcaagaa nnnaagtgga gggtntngg nnnncccggt gaagccctnt 60
ggcaacgga nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnntntntnnn nnnnnnnnn 120
nnnnnnnn nnnnnnnnn nntcaggtgt ccaaaatcct caaannncacn gnnnnnncaag nnnnnnnnn 180
nnnntataan nnnnnnnnn nnnntgacag ataaggcaac eqaatcgagtt aaatctactnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnt ttccttttaa aqmgnnnnnnnc tgnnnnnnn 300
nnttttataa gaagtttttt tttacataa aatataag aatgagcgcg aagaaataga 360
accaagtc agccatctttag tgagacat gttgagttcat tttagcaca aaggggaatta 420
aagacgagc agagaaaaatc caaagttcgc gaaataacgc cttagagttg cgcaaaactc 480
aaaata 486

<210> SEQ ID NO 155
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-296
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 155

aatttatctt ttttccccag cnnnggtaga ggganntctga nnnccctttt gaagcnnnc 60
tagcagctag acnnnnnnnn nnnnnnnnn nnnnnnnnn nnatatcnn nnnnnnnnn 120
nnnnnnnn nnnnnnnn gtnnaggggt gczeannntct xnnnnnnncag gnnnnnnnn 180
nnntattatn nnnnnnnnn nnttgacag atcagagcag aagttgagtt aatnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnn cccttccctt ttcgtnccgc xtnnnntcgc 300
aaaatagaga ggggcttttt atatgacag tttttggaga gaaattcaag aaggggaatta 360
aattttgctac caagcgctat ctttttaatc cagatcgcgt tnttgatgca catccagata 420
aattttgcag tccaatattc tgcagcattt tagttcagat xttttcaca gattccgcag 480
cgcctg 486

<210> SEQ ID NO 156
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 156

taattgtctt ttaaatgag tnnnggtaga ggganntctg nnnccctgtt gaaaccnnnc 60
ggcaaccttt cnaannnnnn nnnnnnnnn nnnnnnnnn nnttccgnnn nnnnnnnnn 120
nnnnnnnn nnnnnnnnnnt tgaagactgt ctaatntctt gnnnnnncga agtgnnnnnn 180
-continued

nnnnntgann nnmmnmnnt gttcagag ataagagaga cttaaaagtg ttcaggttat 240
ttggtatcg aaeccaaac aennnnmcc tctctagnnn nnmmnnmnn tctnnnnnnnnnn
nnnnntcatg ggaggttttt tattggcaca aaatagagag gataagttga taggtatgttt 360
aaggggctatt gttccaaact tggggtatccc gacacttggg ggggaaacgtg aatggaacgct 420
tgtgtagga aatagttcga atagggtgct ttcgggaagag gataagttgg cttgaaacga 490
ggctct 486

<210> SEQ ID NO 157
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 157

tgtgagacat ttatacgag tnnmggtgga gggannaatg nnmmccctat gaagccmnmc 60
agcaacctaa acacacacnn nnmmnnnnnn nnmmnnnnnn nnnntccann nnmmnnnnnn
nnnnmmmmnn nnmmttattgt ttttaagggtg ctaagntcct gnnnnncnncc aacaamnnnn 120
nnmmctanmn nnmmnnnnnt tttttactagag atgagaaggga aagttgctca tttgaaaaaa 240
tgttcttmmmmmm mmnnmmmmn cttttttctgt tnnnnnnnnnn nnmmnnnnnnnn 300
nnnnnagcga aaggttttttt ttggtatatca gaagtttacaa aaggtgtattag agagattttc 360
gttccaaaac gttgatttac aaatccctgtc ccagaaactc aaagttttgct cagttcatca 420
tgtcgattta gaaatgtagc aagggcagat ttgggaggtt gtaagttatt cggagctgg 490
taaag 486

<210> SEQ ID NO 158
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 158

ttacaaatctg ttatccagag tnnmggtgga gggaaatcgg nnmmccccagt gaacccmnmc 60
ggcgcgcag cnnnnnnnnnn nnmmnnnnnn nnnnccaannnn nnmmnnnnnnnnn 120
nnnnmmmmnn nnnnnnnnn nnnnctttagc ctaatnccg annnnnnncc aannnnnnnnnnn 180
nnngtatantan nnnnnnnnnnn ctttctgcag ataagtagta gtttttaatg aggnmmnnnnnnnnn 240
nnnmmmmmmmmnnnnnnnnnnncg otggcttttcccccc cnnnnnnncc aaaaaaaanmmn 300
nnnnnagcagag aagggttatttt tttggtattt aagaggggta gttttgtttta gttgaaagaaaa 360
ntttttatttag taaggtttata ccggttttttc gttttgttct ttaacgctttc ccggaggtctct 420
ggcggagttt cacacaagc aagggtttca gggcacaagc aagagggggg cttottaatt 490
atcgggt 486

<210> SEQ ID NO 159
<211> LENGTH: 486
<210> SEQ ID NO 160
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE: misc_feature
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-303
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 160

tagtttttt ttatcaagaa mnaggtgga gggannctgg nnmoccttt gaagocntntn 60
agacaagga amnnnnnnnn nnnnnnnnn nntttattnn nnnnnnnnn 120
nnnnnnnn nnnnnnnnn nnttacggtt etattnnca gnnnnnnnecag nnnlnnnnn 180
nnnttatatt nnnnnnnnn nnntggaag ataagctggga aatccaaatt taggaaactt 240
ttatnmmnn nnnnnnnnn nnnnnnncct tctctgcggg nnnnnnntt atacnmmnnnn 300
nnntcggctg gaggtttttt ttggagaaat tctctgataaa tacatattaa agagaggtg 360
atttttaga taattgaaa aatcctggg aacagcgccg aacacccggg 420
aagagagat gcctcttgaa gccccaaaattttctttt ccctttttg 480

gcccag 486

<210> SEQ ID NO 161
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE: misc_feature
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 161

acaagtaaac ttatcaagaa mnaggtgga gggtnctgg nnncccaagt gaagocntnt 60
ggcaacgga mnnnnnn nn nnnnnnn nnnnnnnnn nnnnttcccnn nnnnnnnnn 120
nnnnnnnn nn nn nnnnnnn nntcaggtgca cccacnncca gnnnnnnnecag nnnnnnnn 180
nnnnnnnnn nnnnnnnnn nncatgacag ataaggcag cgaacagct aatcactt 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnet ttcctttaaa agmnmmmmnc tgnmmmmmm 300
ncttttgggg gaagtttttt tttgtaataa aataatacg aattyaggtc aagaaaaatga 360
atcaagtggc acctattttt gcagatcagtt tttgaaatctttt cagagcaatta agaatta 420
aagagagcgc gyyagaaaaaa caaagttggcg aaattacac tcaagaatta cgtaaaattg 480
aaatag 496

<210> SEQ ID NO 162
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE:
<221> NAME/KEY: mico_feature
<222> LOCATION: 22-295
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 162
aattatatct ttatcagag ctnnngtga ggganctga nnnoccttt gaagcccnnnc 60
agcacaacctc anmmnmnnnn nnnnnnnnn nnnnnnnnn nnnatasaan nnnnnnnnn 120
nnnnnnnn nnnnnnnnn ntgaaaggtg ctaannntct gnnnntcag gagnnnnnnnn 180
nnnnaatn nnnnnnnnn ctctgtaaag atggagccaa aggtataatt atannnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnmmgc ctttctctat ttegtcgegcn tttntctgcg 300
aasatagaga gaggcttttt atagtgaacgc tatggaggag gaaattaang ggaaattaaas 360
aattggcata aaacgctgac ctatttcat gcgaatcgggc tttgatgaga catcagata 420
aasatagcga tcaatatatt tggatgcatt ttagatccat tatcctcaga aatacgaga 480
cacgtg 496

<210> SEQ ID NO 163
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
<220> FEATURE:
<221> NAME/KEY: mico_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 163
taaattatatc ttattatagc tnnnngtga ggganctgg nnnncccttt gaacccnnnc 60
agcacaacctt caannnnnn nnnnnnnnn nnnnnnnnn nnttccmnyn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnntn ttgaagagtc ctaatntct gnnnnnnnn cagtgmmnnnn 180
nnnntgann nnnnnnnnnt gttcagag ataagagaca ttaaaaggt ttcactgtat 240
tttgtatagc aacctccaa annnnnnncc tctct-agmn nnnnnnnnt tctnnnnnnnn 300
nnnntgncteg gggggtttttt tattggcaca aaatgtagag gataaggtc taggtatgtg 360
aagagcggatt aagttcacaat tggggtatcgc ggaactttgg gagaactcttg aattggcaca 420
tgtgctagaaa aagtttttgaa atgggtcgctg ttcagagag gaaattatgg cggaaaaaca 480
agcctt 496

<210> SEQ ID NO 164
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Listeria innocua
-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 164

tgtagaacat ttagcagag tnnmggtgga ggganatg nnnccctgt gaacccnnnc 60
agcaacacta acataaann nnnnnnnnn nnnnnnnnn nnnnttannn nnnnnnnnn 120
nnnnnnnn nnnntaatgt gtttaaggtgt cttaagtncat gnnnnncag aacaannnnn 180
nnnnagatnn nnnnnnnnn tttttcagag atgagaagga agttgceca tttgaaaaaa 240
tgtttnnnnn nnnnnnnnn nnnnnnncct ctttcagcnn nnnnnnnnnc atnnnnnnnn 300
nnnngagcaag aaggtttttt tgtatatcag aatgtagaag aggtgataga gatgatcag 360
ttaagagaag tgtaataaaga atatactgcccc agaataaaca aagtttcctg acgtcgccat 420
gttgttttag aatattgaac aagttgagatt ttcggatgtag tgtgtatttc aggggctgt 486
aasagt

<210> SEQ ID NO 165
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 165

tccatatctc tattgtagc nnnaaagtga ggagacnttg cgcccttgct gatctactnncc 60
agcagcagca tnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnttatnn nnnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nnnagcagcgtg ctaaanccca annnnnncga gnnnnnnnnnn 180
nnnnnttann nnnnnnnnn nnnntcgaaag atagatataaa agannnnnnn nnnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnccttt cttactctttt nnnnnnnnnn nnnntccnnnnn 300
nnnnncgggtg aagagtcttttt tgtgtaagg agggagagac aatgaccaaat taccagag 360
atactttaaa cttctgggaa tttatcagc aatctgagccga agttgataga atactcgcgt 420
tgagatatga gcatgctggt gtagctggaac acacataggt tgtgtttgtg catctttaaa 486
cctgcca

<210> SEQ ID NO 166
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-300
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 166

cgcgratctc ttagcagag tnnmggtgga ggganatgtg nnncccttcac gaacccnnnc 60
ggagacgcgt tttttttattn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnntatatnn 120
nnnnnnnnn nnnnnnnnn nnnntaatgt gccaatncac nnnnnntaagagttnnnnnnn 180
nnnnntttann nnnnnnnnn nnnntttctgat gttgagagaa caatctact 240
nnnnnnnnn nnnnnnnnn nnnnnntng ctttttctaat tttnnnnttc atnnnnnnnnn 300
gatattgaga aagcattttt tatattatta agcaacacag ggaggaatca aagtgtattga 360
atattaaag gcgttagatg aatatacag ctaaaatataa gaagctcgtg cttgtagtca 420
cgtaattta tagattcag cagatagat ttaatggcgtc atatgttatattggtcttcttcttgagcagg 480
aaaaag 486

<210> SEQ ID NO 167
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 167
aaggttttct ttatctgtag tnnmggttga gggactatg gnnacccac gaaacacnnmc 60
agcaacattc cttttttttt nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnmmnas agaagaggtg ccaanncgg ttmttgctac annnnnnnnnn 180
nnnnaatagn nnnnnnnnn ngctgcaagc ataagagca agtgagctat tannnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnng cctttcctct nnnnnnnnnn ttnnnnnnn 300
natsagttag aagtttcttt tattttagct acagagagcg aatitcctga ataataaatt 360
aagagagcc caactgctatc gttactaacaagc gttattactc tcagactctg ttctagaggg 420
aacttcagac aaaaagcttg acacattgctc agatgaata tttagagcata tttaaaaaa 480
cgacc 486

<210> SEQ ID NO 168
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-302
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 168
taagcttac gtttctgag nnnagtggta gggatcttg gnnnocctat gaagcttmc 60
gggacactn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnntcagann nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nnnnmatgtg ccaaattccaa gnnnnntaaa cegnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn ttggtttgaag ataaagcagtt aaagacactg aaannnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnngttgagaa agatgatttt ttaattggaa aggagtttaa aagagttggaa gttacataaa 360
agagcaagtgc tgggcatatt tttccccctgt ttttttttgc ggctggttt ttggctgtag 420
gtatttacag aggtgatttt actcttacatg caattaaatg tgaacactga ataaggtaa 480
tttgttg 486

<210> SEQ ID NO 169
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEY: misc_feature

<400> SEQUENCE: 169
ttcatacgcgctcaatcagagnnnnncagagggtaaaaaagcagaaagcagacgacggcagccactgcgtggtacttacgtggtacagtagctgcagactgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctgtcctgctgctgtgtctg
cttcgtagaactgcaagtgcacactgtgaagggcatcctgttgaagaatactcttgagttcagacctgtgctaggtgccactagc
486

<210> SEQ ID NO 172
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 172

cttcgtagtc agtcactgaa ggacacctgt acaagatctgt tgaaccagatt tcagatgcca
420
tttggatga aatctttaaa aagaaccttt acgcocgcgt ggcagttgag acaagtgttaa
480
catcag
486

<210> SEQ ID NO 173
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-299
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 173

ttcacatcct ttaatacagag annnngtgga ggagatctgg nnnmcccgat gaaaccmnnn
60
ggcacacgc ommnmmmmm nnmmnmnmnm nnmmnmnmnm nnmmnmnmnm mmnnmmmmmm
120
nnmmnmnmnm nmnnnnmnngt ggggatgctgg ccaatntcct gnnnmmmmmc gnnnnnnnnnnm
180
nnmggttattnn nmnnnnmmnnnc ttcctcctttt cttttttattttt tttttttttttttt
240
nnnnnmnmnmnmnmnmnmnmnmnmnngt cttctctcct cttctctctct tttttttttttt
300
nnnnnmmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmnmg
360
aagagcggct cagtaattttt cttgaggtgc aaagagggag taaaggtggtg tgggtggga
420
ccttgccactg tgggacagag aatatatatc ataqagagag atacaagatgagagattgttgcttgggccaga
480
tgctct
486

<210> SEQ ID NO 174
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Fusobacterium nucleatum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-307
<223> OTHER INFORMATION: n = g, a, c or t/u
tgygaaataaa ccatacaag ccnnagattga gggannccag gnnnnccggt cagatctnnnc 60
agcaacctac gnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 120
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 360
gaaaggaattt acataacctta cccatacaag cccttccacca gggacatccg ataaatcttc 420
agcaataa atccagtagga ttattatgct ttgttttaaa gatgacccota atccagagt 480
tgcctg 486

<210> SEQ ID NO 175
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Pseudomonas nucleatum
<220> FEATURE:
<222> NAME/KEY: misc_feature
<222> LOCATION: 21-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 175
asataasaa ccatacaag ccnnacgga ggganctgg gnnnncccaat gatggttnmc 60
agcaacctac gnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 120
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 300
nnnnnnnggt atggatttttt tattaagta agaatattt atagaagaag gggataaaa 360
tgtattagact tggaaagta ataaaaattt attccaat tttgcagctg gtgaaagatg 420
tatattaata agttatgtg gggataatct ttggaattat aagtttaagt ggtgtggaa 480
aacctt 486

<210> SEQ ID NO 176
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Deinococcus radiodurans
<220> FEATURE:
<222> NAME/KEY: misc_feature
<222> LOCATION: 22-266
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 176
aggtgcacct ttatccagag tnnccgccga tggacntgag gnnnccctg nnnccggnnc 60
agcaacacg cnnnnnnnnn nnnnnnnnn nnnnnnnnn nttcatcact nnnnnnnnn 120
nnnnnnnn nnnnnnnnn ggcagcggtg ctnmttccca gnnnncncc gcgcagcagc 180
cgcgcagca ggggagcag ggggggagac ataaaggaag gcggctctcc tgtcgcgggt 240
ccaacggagctgtcagccnn nnnnnntng gcgcgctccct tccaagcctc ttcttgcctca 300
gggagtggaac ggcgggtttg ggggcacgct tgggtctcc ctccgagggc cgcggccgct 360
gaccttccgctttcccc gacgctttgca ttggaagggc gtcagaaaaa cttacccogg 420
ccagcgccg ccgctggaa ggcattttg cctcaaggtt ggcctgcccc gcggcaggg 480
ccgactat
486

<210> SEQ ID NO 177
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Deinococcus radiodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-315
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 177
ccgagccgcc tctccagct gtnccgccc ggtggattt cgtncgcccc taeccggcnn 60	agcgcaggg ccnnnnnnnn nnmnnnnnnn nnnnmmnnnn nntctcagcn nnnnmmnnnn 120
nnnnnnnn nnnnnnnnnn gggctcggct gnnnntnccg gaaangggg coggttagg 180
ggcgcagca tcggcgcagn cgcccccnnn atgccccggc gggaggtctg ctctcaaccct 240
ggcgcagc acggctggtt ggtgccccmn nnnnmmnnnn nnnnmmnnnn 300
nnnnnnnn nnnnngccaa tcggcctcag aacccatcaac gttcccaagc taecgagggcc 360
ggcgcaggg gcgatccgat tcggcccagc gGCgtggtggta gcgacttcgc gcggccac 420
ctcaagcgg cggctttcgc cttgggacga gcgcaccccc cgcggaggtc tagggcgcaac 480
tctcgac
486

<210> SEQ ID NO 178
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Xanthomomas axanopoides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-315
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 178
cctcggctcc cctcggcaca nnnnggcggc gggancagg nnnnnncctt gatgcogmg 60
nggcgcaggg cggcgcngnn nnnnnnnnn nnnnnnnnn nnnnncacnnn nnnnnnnnn 120
nnnnnnnn nnnnngcggcc gcgttggctg ccaatncc gnmmmmmg ggammnnnn 180
nnnctccggn nnnnnnnngt cggcggagaag aggctctgg cggccgctgg 240
acgcggagtc cggcagaagt cgatggcnnn nnnnnnmmnn nnnnnmmnnn 300
nnnnnnnn nnnmgcgtcc acctgtgata cgcagatgac gctcggatg aacctgcagc 360
gctgcaccc gatttccgt gaccgcccgc gcggcagcgc gacggcacac actgcggctg 420
gccgacgct tgtcagcgc gtcgagtgc ggcgtcggcg cgtgcgacag ctggcgtgc
480
gtctag
486

<210> SEQ ID NO 179
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Xanthomomas campestris
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-315
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 179
cgtacctca ccctgagac nnnegcggag ggannccagy nnnncctttt gatgcctgng 60
ggagcggcag gcgcggygmn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnn nnnncggoec gcgttctggtc ccaaatnncet gnnnnnngcg gaaannnnnn 180
nnnnntccgyn nnnnnnnngt cccgagcgag aatgtgctgaa tctgtgctcct tggacgcgta 240
acgcagctc cgcgaagct cggattgcnc nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnntggacct accccgagata tgcocatagc ctctctgacg accacagtgc 360
cacatcaca acgcagacac tacacgcccgc ccgagtcgag cagacgcgc gcccctcgtgc 420
gccggcgagt cgtcatcaac ctaacggag ggccggccgg ccaacgccag ctgcegtctc 480
gtaag 486

<210> SEQ ID NO: 180
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 180

ttacotaacc ttattttgag nnnnagctga gggatntttg gnnccccata gaagcttmcn 60
agacacccag ccnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnntttnnn ccacnnnnnnn 120
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccacnnnnnnn cccnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
300
nnnnnnggtta ggcacctttnn tattttcag agggacaaacc attttggacct gatattgacg 360
ttactacatt agacatctggt gattttttaaa ottacttctgg tggacagatt gataattttac 420
gtctcagctga tgaacatctga ggaacctcctg tgcacccctct tctcgttgtg tggcatcgcac 480
ttactg 486

<210> SEQ ID NO: 181
<211> LENGTH: 466
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-486
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 181

acggatcctc ttatcttgag nnnnggttga gggacatagt nnnccccaaa gaacccnnnc 60
agccaccttc tnnnnnnnn mnnnnnnnn nnnnnn mnnnnnnnn mnnattnnnn nnnnnnnnn 120
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
180
nnnnnntatgt nnnnnnnnn nnnntgctgacg atgacgctttt gacgccttc 240
tcctcttactc tnnnnnnnn mnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
480
<210> SEQ ID NO 182
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Geobacter sulfurreducens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-303
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 182

gtagaccttc tatacaagag mntngtggaagg mnnccctgt gaasccancn 60
agcaacogt cgnmnmmmmn nnnnnnnnn nnnntgtnnn nnnnnnnnn nnnnnnn 120
nmmmmmmmm mmmnmnncgg agcgcaggtc taatanccct gnnnnnncc nnnnnnnnn 180
nmmgaaannn nmmmmmmmm nmsgsagcg atsgagggga tttgtgacc acgcgcggt 240
acnmnnnnnn nmmmmmmmm nmmnmnngg ccocctcccg gnnnnnnnt ttcnnnnnn 300
nnncggcagg gggcccttcct ttttcgcccg cgcgcgcaac gcgcgcgggg gatactgtc 360
cgtgcgcac tgcgaagaac aatcgtcaca cttccgaacag gatctcaggc tggaagccgg 420
cggagatcgg ggccacatca ccocgtgcca cgcagactac ggccggctga aacgcgccg 480
gtccaa

<210> SEQ ID NO 183
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Geobacter sulfurreducens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 183

agcgcacata ctatacaagag mnnogacgga ggannccagg nnnncoggggc acgctcgncn 60
aggcaacttc ccmnnnnnn nmmmmmmnn nmmmmmmnn nmmatgmmn nmmmmmmnn 120
nmmmmmmmm mmnnnmmgg gggasagggt ccataatnccct gnnnnnnccg gacommnnnn 180
nmmggacann nmmmmmmnng gtttccggag ataaaaggaga ggcagaccce tcaaggtgaa 240
tegaannnnnn nmmmmmmmm nmmnmmntc ttctccgcmn nmmnnmmmcg acemmnmmn 300
nmmncggas ggaggatcttt cattggccag gaaaccgta acatccgcac gcagggcgcga 360
cgactgctgc tgcaggtgcca tatccgcacc gggccgggtg cggactcccc cttccggacg 420
gcagactccgg gcagctggag attgggccag agcaagggcc aagattacttc cgcgctccgg 480
aaccoc

<210> SEQ ID NO 184
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 184

acacatacttc ttatacaagag tnnnggcgga ggannccggg nnnncogcgat gatggcnnnc 60
ggaacagcag ccattgmnmn nnnnnnnnnn nnnnnnnnn nnnnaecnmm nnnnnnnnn 120
nnnnnnnn nnnnnntata agctaaagttg ctattncct gnnnnncaca atgammnnnn 180
nnngtttmm nnnnnntnc tccttttgaag atasagaggg atctattttccttcttctg 240
nnnnnnnn nnnnnnnnm nnnnnnncgc accttncmm nnnnnnttta ttttttttttttnn 300
nnnnnnnnn nnnnngatgg ggtgttttttt atttttatgac atatatgag gcgaaactat agatgaaaaa 360
agttattta agctagttcag gggacgcgtg agttattta ggcctgctgaa ggcctgctc 420
gggataagaa gagaaagcgt tagatagaa aaagattact gctggttga cagggcgggc 480
gcatga 486

<210> SEQ ID NO 185
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-303
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 185
agcaatttaac ttatcagag mnnaagtagga gggancttg gnnncocctat gacacctnnnc 60
agcagcgggt tcctttttmm mnnnnnnnnn nnnnnnnnn nnnntatann nnnnnnnnn 120
nnnnnnnn nnnnnnnnm nnnnnnnngc gagacccgtg ctattccca gnnnnncaca gnnnnnnnnn 180
nnnnnnnnn nnnnnnnnnn nnnnnttttaa nnnnnntgaaag ataatgtatg ggcctttcctt tattaannnn 240
nnnnnnnn nnnnnnnnnn nnnnnnnnc cttttatcta nnnnnnnttttttttt ttttttttttt 300
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnntatggc aaggttttttt gttattcaca aagagaaaaa gagaatagg gaaaaagtacg 360
ttttataac ccattttaaatt caggttttta gggggttttg gaaggtatatg taatctaaaaaa 420
atatcaggtt atgtgttctca actatagggt tatttacatcagctgacat ctaatatttctttaa 480
tattat 486

<210> SEQ ID NO 186
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 186
tttatccttttt gttatcaagac mnaggtgga gggancttg gnnncoccttt gaaacctnnnc 60
ggcagcagcag tcaacmmmnnn nnnnnnnnmn nnnnntttttnt nnnnnnnnnn 120
nnnnnnnn nnnnnnnntn gataactgct ccaaccttnc gcnnnnncaca ntnnnnnnnn 180
nnnnntattnn nnnnnnnnnn agcttgaag atasatgtag ggacctcggtt tatatacagg 240
tgtatttaaatttctga aaaaannnnntt ccttatttttct nnnnnnnn ntaanncnn 300
nnnnngagaaag ggaggattttt tatatttttctt ttctttatcagctgccaa ctaatattttattttt 360
tagggagaga atcaatggaa aagagatgg ggagggcgtg tgtatcgagtt tgtgaggga 420
agtttttttaa taactttgttg ccggttttttaga aaaaacgga cagccgagc aatgccaaaa 480
gacgag 486
<210> SEQ ID NO: 187
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-298
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 187

cgatacattc ttaatcaggag nnnaggtgga gggannctgg nnnnccctac gatacctnnc 60
agcaacctgg cttnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnnatasaa nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
taagagagg agcactcttttt ttttcacctc gggagctcta ctctaagttt ttacagcata 360
tagagagggg aaaaatgatt tcttttata atgtaatgaa atgatatgaa tacaggtggc 420
aatctgcttc tgcgggtgga gatgtaacgt ttaagtgtga gaaaggggaa attttggcca 480
ttatcg

<210> SEQ ID NO: 188
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 188

gataatctt ctttaatcaggag nnmgggcaga gggannccgg nnnncccttt gaagccmmnc 60
agcaacctca gtcctnnnnc nnnnnnnnn nnnnnnnnn nnnnatacnnnn nnnnnnnnnn 120
nnnntttttnn nnnnnnnnn nnnnnnnaac tgaatagttg cttatnntct gnnnnnnnca aatgcnmmnn 180
nnnnmttttnnn nnnnnnnnmgc attttggas aaaaacgta actatttgg taaaananmmnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnct aatnntttcm nnnnnnttg aatcnnnnnnnn 300
nnnnnggaag gttagtttttt tttatattc aaaaatatat aatggagtttta taaaatgaaaaaa 360
gttatgcac cttcaacaaac atggaaasat aataatgtactttaattttctgg aacacaaaaa 420
ataatcattag aagcccttca aaggggttga gaaacaggtg atcaagttac agatttccat 480
tctgct

<210> SEQ ID NO: 189
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 189

aatcaaagc tattcaagag nnnagcgga gggannctgg nnnnccctgc gaagctnnnc 60
gggacccctgc tttttnnnnn nnnnnnnnn nnnnnnnnn nnnnataagann nnnnnnnnn 120
tgaacacttc tttataagag  nnnagcgga ggganctgg ggnnoccta  cgtgcctmncc  

60

ggcagcggac tcnnnnnnnn nnnnnnnnn nnnnnnnnn ngattttttaa  nnnnnnnnnn 

120

nnnnnnnnn nnnnnnnnn gtagcgctggt ccaatnccaa gnnnnnnnccaa gnnnnnnnnn 

180

nnnnatgnmn nnnnnnnmnn  ngctgggaag atgggaagag cgttcctttaa agatgtagaa 

240

nnnnnnnnn nnnnnnnnnn nnnnnnnnn agcctctnnn nnnnnnnnn cccttttctnnn 

300

nnnnnngggg acctcctttt ttattcatta gaaaaaggt tgcacagtgg gaggaggttt 

360

acctggaag  aaagcagggg aatggtgctt gctttattac caattttgggat atttttgggct 

420

tttttgaatt ccccttttgaa attctttcagt gttctttcctgaa atttctttgcgtt 

480

attc 

486

aataattac ttacccggag nnnaggtgga gggacnggnm nnnnccctat  gagaacctnncc  

60

agcaacccct acgtnnnnnn nnnnnncccc nnnnnncccc nnnnnnnnn nnnnnnnnn 

120

nnnnnnnnnn nnnnnnnngc ttaagaggtg ctaaatnccgg  nnnnnnnccag agacacnnn 

190

nnnnntrtnnnnn nnnnngctgtg ttttgggaag atggagaggt tttggaagct  gaagaaan 

240

nnnnnnnnnnnnnnnnncgc accttcttmmnnnnnnnncgttgtnnnnnn 

300

nnnnnmmaga ggtctcctttt tgggtttagaa aagggagctg  cctgtctgata atctcttttc 

360

aaaaataatt aagctgataa aagttcctactaattacagag gggattataa atgacacgatt 

420

tatccacaaaaa attagagtaa gcatacggaa ttggatccga attatacggg attataggac 

480

ttggg 

486
<210> SEQ ID NO 192
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 192

atgaaatcc ttatacagag nnnaggtgga ggganctgg gnnnccctat gaaacctnnnc
          60
ggcagcggat tggmnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn
          120
nnnnnnnnnnn nnnnnnnmt gaaatctggt ccaattncga gnnnnnnncaa gnnnnnnnnn
          180
nnnnngtaannn nnnnnnnnnn nnncttgaaag ataaqaaaaga agoctaatttt gactatatatat
          240
acagaamnnnn nnnnnnnnnnnc cttcttcttan nnnnnnnnnnt cttttttnnnnn
          300
nnnntagaaa gaggcttttt taeqgtgaaaa taaaaggaag aagaaasaatgg ggagacacag
          360
gagtagcgc acaaaggaaa acaatgtgaa agaatctgga aagaataaaag gaaaaagataca
          420
tagaaaaacag tgtatgattt cagcgaatcc cggagatggtg ttaactaaga taattacgcat
          480
cgatag
          486

<210> SEQ ID NO 193
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 193

gaatatcttc ttatacagag aannnggtgga ggganctgg gnnncctgat gaaacnnncag
          60
agcaacogen nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnnnn
          120
nnnnnnnnnnn nnnnnnnmt gtttgaggag ataaagacga aqatatactag taamnnnnnnn
          180
nnnnnnnnnnnnn nnnnnnnnc ttttctnnnnn nnnnnnttatctnnnnn
          240
nnnnnnnnnn nnnnnnnncct tttctnnnnn nnnnnnttatctnnnnn
          300
nnnnnnnnng agaggtttttt ttatgcaaaa aaaaaaagatt gaaaaaaaat ttatattaag
          360
aagaaagggg ttgaaagct cttgacact cgaabaatc gtaaactcgt gtaattacagt
          420
ttatgtaactat atgatagcag aagataacac aatattcattg ttagatatc aagaaacatat
          480
cgttct
          486

<210> SEQ ID NO 194
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 194

tatacaactc ttatacagag cangnggtgga gggatntgg gnnncccgat gaagccnnnc
          60
agcaacgccc cmmnnnnnnn nnnngttaa taccctggtg aatggggcgct ttttagcgcc
          120
casaaaaannnn nnnnnnnnn nnggacgggtg ctaattncga gnnnnncag aagtnnnnn
          180
nnnnnnnac ttctggccag ataggggg agaagataaa cttcaamnn 240
nnnnnnn nnnnnnn nnnnnncc ttcttctnn nnnnnnnnt agtnnnnnnn 300
nnnnnnggaa agaggttttt ctaagtcga aaaaaccttg aatgaaaaaa gggggagaag 360
cagatgggt atatattcatt aacagaagta accgtgtac aatagtcgaa agacagtgtt 420
tattttgaaa agaagcgaag tcttttttgt ctcgaatttg gaggcggaa atttattat 480
gtgttc 486

<210> SEQ ID NO 195
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KAS: misc_feature
<222> LOCATION: 23-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 195

taaactttc ttatcagac cagggtgga ggganccag nnnccgcac gaaaccnmc 60
ggcaacgta ctaacannnn nnnnnnnnn nnnnnnnnn nnnntaamnn nnnnnnnnn 120
nnnnnnnn nnnnnntgtag acacagttgc taatntntc tnnnnnnncc cnnnnnnnn 180
nnmattam nnnnnnmmn nnnntcgac ataggggtct gctttgaaac aaaaaanmnn 240
nnnnnnnn nnnnnnnnnn nnnnnnmcct tctnnnnnn nnnnnnnncnt tegotnnnn 300
nnnnnnnnn agaggttttt tttattaaact aggaggttat aacaatgagc ggaattag 360
cgacgattt aacctgat gattcagac aactagaaaaa aaaaagtgcag caatgcaac 420
tctgtttca aaattgcctt tggactccat tcggccactt attgcaaggag cagtttaagc 480
agcata 486

<210> SEQ ID NO 196
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KAS: misc_feature
<222> LOCATION: 21-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 196

agcacaacatt tctcttagc nnnaggtcag gggancttg nnnccctat gacgcctnmc 60
agcacaacatt acnnnnnnnn nnnnnnnnn nnnntattnnn nnnnnnnnn 120
nnnnnnnn nnnnnnntgta taataggct ctaacctaca gnnnnnnnca aattnnnnnn 180
nnnnggaan nnnnnnnnnn aattcag acagagaagaa gcctctttac aaaccgaan 240
nnnnnnnnn nnnnnnmcct ttctnnnnnn nnnnnnnncnt ttnnnnnn 300
nnnnnnnnn nnnnnnmcct ttctnnnnnn nnnnnnnncnt tegotnnnn 360
nnnnnnnnn aatattttatttttttttattaatttc gtttcatttttt aaaaaggg 420
aatacttttac aatgtacatg ccggccaaaa tctcgtggcca tcgagccacg gacctgaact 480
accaaaaggr ctttttaaact gctttttttc tttttaatctt gttgtagttaaa cagaggaatt 486
ggttag 486

<210> SEQ ID NO 197
<211> LENGTH: 466
<210> SEQ ID NO: 198
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 197-210
<223> OTHER INFORMATION: n = g, a, c or t/u

<220> SEQ ID NO: 199
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-323
<223> OTHER INFORMATION: n = g, a, c or t/u

---

aagacacact ttattgagag cmmuggtgga gggannaggg nnnncctgt gaaacgmmnc 60
ggcaaccttc aacncnnncn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnnn nnnnnnngtt tga accggttg cta tancct gmmnnnnncc aaccnnnnnn 180
nnnnngaatnn nnnnnnnnnn gttttgctat ataagggagg gaaaccattat gttnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnncc ccttccan cnnnnnnnnn anagnnnnnnnnn 300
nnnntgagaga ggggggttttt tattagtag agaatgaggga gattttggaa attactagat 360
tatgtgcaaa aaggaattgt agaatggtgt ggtggggttg gaaaccatttt acatcaccac 420
ggttgtgcaaa ttaggttttga aagatagat ataatcgtc tagattttaat tatacatt 480
cataag 486

---

aagacacact ttatccgag ctmnggctga gaggannccgg nnnncocctag gaaacgmmnc 60
agcaacactt cttgtamnnn nnnnnnnnn nnnnnnnnn ngtggtaaan nnnnnnnnnn 120
nnnnnnnnn nnnntacaggg tgaataggg gtaaaactgc gnnntgnega ggtcnnnnnn 180
nnnnnacann nnnnnnnmmng cttcgggaacg ataaagacgg aggccaaara gcaatagca 240
agtagaataat taaaannnnn nnnnnnnnc tttcctttnnn nnnnnnnmat ataannnnnn 300
nnnmcagag asmgtttctt ctgtattcctt gttgggsgga ataatttga atgcacactc 360
gttggcagatt cttggttagc tocttagcatt tatacaatt actgtaggg ggtttaccac 420
tagacaaaaa aagccatcctt gttccacatt ggttgcggaa cttgaaaaa tccagaaaaa 480
atgtt 486

---

tctgattttctc ttatacaagag annggtgtga gggacntgtg nnnncctgtg gaaacgmmnc 60
ggcacgcgtc aacncnnncn nnnnnnnnn nnnnnnnnn nnnnttatnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnngtt tgaataggtgc ccattnncct gnnnnnnccaa agcnnnnnnnn 180
nnnnaatgn nnnnnnnnn nctttgtagg atgagagagaa gggaatgct tgattataac 240
-continued

gcatataaannnnnnnnnnnnnnnnnc ttctgtcnn nnnnnnnnnnc tetannnnnn 300
nnnnaacgg yaaagtttttt ttggtgttttg aatggtgagg acattcaaat aataaaagta 360
atgagaacgg tgggctaacg tatcaaaat aaaaatgdc ggagtcaatc aaaaatctga 420
cctcgacgcc tgaacacgtc ggctgttctca tccctctccat tcggcaaaa agcccctctca 480
agctctg 496

<210> SEQ ID NO 200
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus anthracis
<220> FEATURE:
<221> NAME/KEY: mics.feature
<222> LOCATION: 22-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 200

ttgcatagtct tatcaagaa annaggtgga ggannncagg nnnnctcgat gaaacctnn 60
ggcacagcc gttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 202

casaacatt ttagtgcag mmnaagtggag gggannccccg nnnnnncctat gaaacttmmnc  60
ggcaacctgg tnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cagacgtg tcactactaag aacatggtg  120
nnnnnnnn nnnnnnnn cagaagagc ccaattntct gnnmmnncag gtgnnnnnnn  180
nnnaagaan nnnnnnnnn cactctgaag ataaagcgcgg ttcaatgtgt caagaagmmnn  240
nnnnnnnn nnnnnnnnn nnnnnnnnc gcctctctatt nnnnnnnnt tcnmmnnnnnt  300
nnnnataaga gtagcttttt ttagtgcctaa aagtttaaggg gggaatggtgt aatggtgtat  360
gtttctttgt tggcgggatttt tggggggtgc ggctggatgt taaattgattgta attctctcgg  420
catcggtttg agtaatggcag aacaacggcg caaagcgccg aacaataggg ttctttaaca  490
acacct  486

<210> SEQ ID NO 203
<211> LENGTH: 466
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 203

aataacaagct ttagaagcgag aaggaacctgg gggannccccg nnnnnncctat gaaacttmmnc  60
ggcaacctgc ttttttttttt nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn  120
nnnnnnnnnn nnnnnnnnn nnnnnnnnn aagacgtg tcactactaag aacatggtg  180
nnnnnaactnn nnnnnnnnn nnnnnnnnnn attttgaag ataaggttaaa atatattatc gaagmmnnnn  240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn mctttctttttmnnnnnnnnn aatggnnnnnnt  300
nnnnnnnnn aagatgtttttttttttt aaaaaggggg ggctggattttttcg ctggtctctct cgagctacgg  360
gaccactttg aggagacgtt gtaaaaaattttaaagagggcgcggcgttctctctc gaatgtatgtg  420
tcaatcataa tttttttttg agatggcaca gtttatataat taagaaaaa tggaaaaatga  490
cggtta  486

<210> SEQ ID NO 204
<211> LENGTH: 466
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-305
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 204

tgaaaccttc ttataaagag mmnaagcggag ggganncttg gnnncctcctc gatgctctmnc  60
gggacgcgcag tnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnngatcttcc  120
nnnnnnnn nnnnnnnnnn ggtgcttctg ccaatctcga gnnmmnncag gennnnnnnnn  180
nnnnataatn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn  240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn  300
nmmnggag aggtcttttgc ttattcatta gaaaaaggtt gaaactaggg agagatgcgta 360
cattggaaga acgagagaga aaggtttgcc cattattacc acttggaata tttttggcgc 420
tattttggt ttctggaatt attacaggtg atttttataaa attgocgata cttgtagaca 480
tttcga 486

<210> SEQ ID NO 205
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 205
aaatattaac ttatccagag nnnaggtgga gggamncgg nnnnoctttat gaaacottnc 60
agcacocct atannnnnnnn nnnnnnnnn nnnnnnnnn nntatatn nnnnnnmmnnn 120
nnnnnnnnnn nnnnnntngt taggaaggtg ctaattncog nnnnnnnncag aagacsmnnn 180
nnnnngattmm nnnnnnnngt ttttggaag ataaagaggt ctctggaagtt gaaagaaaan 240
nnnnnnnnnn nnnnnnnngt accttcttnn nnnnnnnnn tggtnnnnnnnn 300
nnnnnnaaga ggtctattttt tgggtgatat aaaaaggaggt gtctgtcatat attccattttcc 360
aaatattaata tagatattaa aaggtgacat attaaggag ggatagtgaga atgataaasat 420
tatcagacaa attagaggtg gccagagaa ttggagcgc gattatacggg attataggac 480
tttcgg 486

<210> SEQ ID NO 206
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 206
atgaaatcc ttatacogag nnnaggtgga gggamncgg nnnnoctttat gatacctnnnc 60
ggagcgggt tagnnnnnnnn nnnnnnnnn nnnnnnnnn nnnntttannnn nnnnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnt nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnnntaann nnnnnnnnn nnnnnnnnn ctaatattg tccacattcga gnnnnnnncaca gnnnnnnnnnnn 240
gcagacnnnnn nnnnnnnnn ctttttttan nnnnnnnnn ctttttttan nnnnnnnnnnnn 300
nnnnntagaga aggctttttt tgtgttaaaa tataagggg gagaasaatg gggagcagagag 360
gagtagcgtc acaaagaaaa acaatttiaa gaaatgttga agagatagag gaaagacgaca 420
tagaaaagataa gcagcattatt gctgcggatt cggagatgg taaccaagag tttacagcact 480
cagaa 486

<210> SEQ ID NO 207
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
LOCATION: 21-305
OTHER INFORMATION: n = g, a, c or t/u

SEQUENCE: 207

attagttttt ttatattaag nnnagatggg ggganctgg nnncccgat gaatactnnn

agcaacgagc tnnnannnnn nnnnnnnnn nnnnnnnnn nnnataaann nnnnnnnnnn

nnnnnnnnn nnnnannnn nagnacgctg ctaagtnccu gnnnncnnac agctnnnnnnn

nnnmatgaan nnnnnnnnn nnnnagagag atgaggggaa atggaatcaac attoanunn

nnnnnnnnnn nnnnnnnnn nnnnnnnncet ctccttatnn nnnnnnnnnn nnnnnnnnn

nnnnnngaag aagagtttttt tattagaga gggggatatg agttgaaatgtgatgtaacgt

atatttttaga aagtttttcag caattattta agtatgtata ctaactttta ggaattacgt

tagttttaat gatttttttt tttgttatag ggttagggtt ggcgttacata acaaaaaaca

aaacga

486

SEQ ID: NO 208
LENGTH: 486
TYPE: DNA
ORGANISM: Bacillus cereus
FEATURE:
NAME/KEY: micr_feature
LOCATION: 22-308
OTHER INFORMATION: n = g, a, c or t/u

SEQUENCE: 208

gaaatttttt ttatcagag annnngtggg ggganctgg nnncccgat gaacacnnnn

agcaacgagc tnnnannnnn nnnnnnnnn nnnnnnnnn nnnngatnnn nnnnnnnnn

nnnnnnnnn nnnnnnnnn nnnncgagctg ctaattnccu gnnnncnnac aacannnnnn

nnnnatttnn nnnnnnnnnt tttctgggag ataaagcagaa gatatatagcc tannnnnnnn

nnnnnnnnnn nnnnnnnnn nnnnnnncet tttctnnnnn nnnnnnnnt tttcnnnnnn

nnnnnnnngg agaggttttt ttattgcaa aacaccagtt acgaaatattt ataattagaa

gaaaggggtt ggcataact gttcactctg aaaaataegt caaacgtgtt agttcaggtt

tagatatat gttgagcagaa gaaagccca tattacggtt agtatattgaa gaacacatcg

486

SEQ ID: NO 209
LENGTH: 486
TYPE: DNA
ORGANISM: Bacillus cereus
FEATURE:
NAME/KEY: micr_feature
LOCATION: 23-309
OTHER INFORMATION: n = g, a, c or t/u

SEQUENCE: 209

taaatctttt ttatcagag canngtggga ggganncag nnncccgag gaacacnnnn

ggcacaagct ctaacnnnnn nnnnnnnnn nnnnnnnnn nnnaatnnn nnnnnnnnn

nnnnnnnnnn nnnnnnngttt agacaacgctg ctaatncttc gnnnnncnag cnnnnnnnn

nnnattaccn nnnnnnnnn nnnngtcag ctaagcagct gtgtgaaa aaannnnnnnn

nnnnnnnnnn nnnnnnnnnn tttctnnnnn nnnnnnnct tattcnnnnn

nnnnnnnng agaggttttt tttatatggtt agaggtttat acaattgagc ggaatttag

360
cgcactattt aatccatgtg gattcacaata acttgaaaaa aaaaactgag caaattgcaac 420
tcggttaaac aatgygcttc tgtgacacctt tgtcacaattt attgcaagaa caattaagc 480
agcata 486

<210> SEQ ID NO 210
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 210

agacaaactc ttataggag ctnnnngttga gggannaaggg nnnncctgt gaaacnmcnc 60
ggacaccttc aascnnnnnn nnnnnnnnn nnnnnnnnn nnnsaatinu nnnnnnnnn 120
nnnnnnnnnn nnnnnngtt tgaaacnggtg ctataancct gnnnnnnnca aacnnnnnn 180
nnnngaaaattt nnnnnnannnn gntttgcata ataaagggag gatcagattat gtnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnncc cctccttcaan nnnnnnnnnn aagnnnnnn 300
nnnnngagga gngggttttt attatgtaga aatnnggga gattttggtgaa atactagat 360
ttatatatc aaggaatgtg aatagggtgt ggtgagggtg ggagcttattt acatcat 420
gggttacaca gtagttttaa ggaatggaat atatctgctc cagatttaat tatactgatt 480
ctaaag 486

<210> SEQ ID NO 211
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-308
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 211

agcacaaaccttt ttatctggag ctnnnnggtag gggannntggt gnnnccttacgt gcggccttmcnc 60
agcacaacttc aascnnnnnn nnnnnnnnn nnnnnnnnn nnnnattttnn nnnnnnnnn 120
nnnnnnnnnn nnnnnngnt gntaattggct ctattnccaa gnnnnnnnccaa aatnnnnnnn 180
nnnnnggaan nnnnnnnnn gattgacac agtgaagaaag gactctattc aaacctgaan 240
nnnnnnnnnn nnnnnnnncc ccttctnnncc nnnnnnnnn cttmnnnnnn 300
nnnnnnnnn nnnnnnnn nnnnnnngag aggggttttt ctttatattc taactaatgc gttcattttaaaaaggaga 360
atattttcttc gctacncatgc gaacaacaat tatgcaaat gggaaaaacgg agtgaacacta 420
cacagggac ggtaaacccctgctttatt tttcaactgc ttatcgctac gagaagattg 480
ctaaat 486

<210> SEQ ID NO 212
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-306
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 212

tatcaactct tatcaagsg cmmnggtgga gggatnttgj tnnnnccgat gaagccnmc 60
agcaacgcac cmnnnnnnn nnnnnngtta tacactttgt aatatttgj cctttttaeg 120
ccaaannnn nnnnnnnnn nggccagtytg tcaattncc caumnnncgag aagttnnnnn 180
nnnnnaann nnnnnnnnaac ttotttgcag ataaaggggg agaagataaa cttaaaaaann 240
nnnnnnnnn nnnnnnnnn nnnnnnnccc ttotttccnn nnnnnnnnt agtnnnnnnn 300
nnnnnnnggag aaggtttttt tcaagcttga aaaaactctg taaataaaaa agggggagaa 360
gacgatgagga tatataagct taacctcaac aacagctata caaatagcga aagaacacgg 420
tatattttga aagaagccaa atgtatttttg tcaatgaaatt ggagatgagaa attaaatta 480
cggtgt 486

<210> SEQ ID NO 213
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEN: misc_feature
<222> LOCATION: 23-307
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 213

ggatacctct ttatccgag ctnmngcgga gggannccag gnnncggat gaagccnmc 60
agcaacotca cttgtttnnn nnnnnnnnn nnnnnnnn attgttatc nnnnnnnnnn 120
nnnnnnnn nnnnacacag tgaatttgtt cttaaancct gnnntgnccga ggtncnnnnnn 180
nnnnnacann nnnnnnnncn gttcctgaag ataaagcag gaagccaaaa gcaggtatcga 240
agttgcaat taaaannnnc nnnnnnnnc ttcctnnnn nnnnnncctt atatatgttnn 300
nnnnnnncgg aaggtttttt ctgtatattt ggtggtggaga attaaatgtat gtgcaaatct 360
gtggcaaat aagggtgat tcctgtacat atataatatt actgtagggga ggtttccacac 420
atgcaaaaa aagctcctat gttoacatct ggtcttgtaa ctgagggaca tcacataaa 480
atttgtt 486

<210> SEQ ID NO 214
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEN: misc_feature
<222> LOCATION: 23-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 214
cgtattttct ctatcaagag anmmngtggga gggacnttctg tnnnnccctgt gaagccnnmc 60
gggcaacotgc aacnnnnnnnn nnnnnnnnn nnntttattn nnnnnnnnnnn 120
nnnnnnnn nnnnnnngt tgaattgtgct ccaatntcc cnnnnnnncaa agcnnnnnnn 180
nnnnnaatnnc nnnnnnnnnncn gtttggagag atagagagaga ggaatagt ggttatatc 240
gacattnan nnnnnnnnn nnnnnnnncn ttctgttttn nnnnnnnnnnctanmnnnn 300
nnnnngccg aaggtttttt tgggtttttgt aatgtggsagac atctatnctct atataaaa 360
gtgcaacgcg gggactaccc gccttaaatc taaaattgtt atgagctgat cccacaaaa 420
-continued-

<210> SEQ ID NO 215
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-301
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 215

ttgcatagct tttctaaga anaggtgga gggannccag nnnncggat gaaacotnnt 60
ggcaacgcc gtnnnnnnn mnmmmnnnn nnnnmmnmm nnnnmmamnn nnnnmmmmnn 120
nnnnmmmmn mnmmnmmnm gggatggt gcnnntctt gnnnmmncag gnnnmmmmnn 180
nnntaaacac nnnnnmmmm nnnccctggag ataaagaaga gocctttggag ctgcttttca 240
aanmmmmnn nnnnnnnnnc nnnnnntct gnttctttct tgmnnnnnnn ttttnnnnnn 300
ncaggaagag gcggcttttt tattttgtat aaagagagag agaataagag atgggagaat 360
catgggggaa aggaaacatgc tgcgtgcaag tggctctacgc ccacaaagat gggtaaccgc 420
gttttttacc ggtttttcag aagcacaacgt ataaatagca taactcggag gatcttacag 480
ccttat 486

<210> SEQ ID NO 216
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-304
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 216

tttaaccttt gtatacagag nnagaggg gggannctgg nnnncccttt gaaacotnnc 60
ggcagacag tcaammmnmnn nnnmmmmnn nnnnmmmmm nnnnmmmmnn mnmmmmmmnn 120
nnnnmmmmn mnmmnnmmnt gaaatctgtg ccaacctctt gnnnmmccaa gtnmmmmmnn 180
nnnntatmm nnnmmmmmmn agctgtagag atagaatagag ggcacctcgtt tatatacgg 240
tgcttaacct gtacgttaaa ammmmmntct cctttcttmm nnnmmntctaatatnmn 300
nnnngagaaa agggatcttttt tattttcttt ttctctcatc aatcatccaa cataatatttt 360
tagggagaa atcacaagaa aaaaaagtgtt gtaacccgtta ttgcatcagtt gtagagatata 420
agtattttc tataatcggag cgctgtagtt aaaaaagaa cagcggcagc aaatgcaaaa 480
ggagag 486

<210> SEQ ID NO 217
<211> LENGTH: 486
<212> TYPE: DNA
<213> ORGANISM: Bacillus cereus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-306
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 217
-continued

tatatatac ttatacaag tnnngccgga gggaanctgg nnnncccgat gatgcnnmc 60
nggcaacgcg cttatannn nnnnmmnmnn nnnnnmnnn nnnncnmnnn nnnnnnnnnn
nnnnnnnnn nnnnntata agttaaggtg ctaatnncct gnnnnnncnac aacngnnnnnn
nnnnntcnn nnnnnnnmc gtnttggaaag ataaagaggg aatctatttt gttctatcgyg
nnnnnmmnnn nnnnnnnnc gccctotcnm nnnnnntta tttttttttttnnnnn 300
nnnmmngaga gcgccttttt atttttaac gtaatattaa gggggaattta tagatgaaga
360
ngattatttt tattattagttttag tattattttat aagcccatgtg agccggagtct
420
cagataaaga agttaaacgg ttgatagaga aaaaattgc tgcgggttta acagggggtc
480
cagctag
486

<210> SEQ ID NO 218
<211> LENGTH: 486
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 21-303
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 218

tagatatttac ttatcagag nnnagtttaga gggaacctgg nnnoccctat gacacctnnn
60
agccagcgggt tcttmmnnnn nnnnnnnnn nnnnnmmnn nnnngtaattnn nnnnnnnnnn
120
nnnmmnnnm nnnnnmmmg gaaaccccg tgttaattncaa gnnnnnnncnn gnnnnnnnnnn
180
nnnccaagtn nnnnnnnmmnn nnncttgaaag ataaagtgtgg gcgcctttgtttt tattaammnnn
240
nnnnmmnnn nnnnnmmnn nnnnnnnnc gcgtatcttaa nnnnnnnnn tttttttttttt
300
nnntaagactc aagctttttttt gcgtatctaa aagagaaag ggagtaatggg aaaaagtacg
360
tttcataaacc tataagattttc tgggttttttttg ggtggtatag taattaaaa
420
attacatttt atggtgttcca cgcctagggttt tattcagacg gtgcatttcc tattatcga
480
tattat
486

<210> SEQ ID NO 219
<211> LENGTH: 505
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 219
	auacuauaugu gggusaacag guuncucucg auucmnmmmn nnnnnngcua nnnnnnnnnnn
60
nnnggggcuug gagcunaagc acggaawenu cgguucguaua cyccmnaae acmnnngcg
120
gagccacgcc gaaagnccg ccgcacugu gancggmnnm nnnnnnnnnn nnnnnnnnnnnn
180
nnnnnnnnn nnnnnnnmmn cagacccuug uccgasauu gnnnnnmmnn nnnnnnnnnn
240
nnnnnnnnnn nnnnnnnmmn nnnnnnnmmm nnnnnnmag ccaugggc
300
nnnnnnnnn nnnnnnnmmn nnnnnncnaa aannnmmnn nnnnnnmnn nnnnnmmn
360
nmmmmggcu cgggaaggc ucgaauagau guugucnmn nnnnnnnnnn nnnnnnnnnn
420
nnnnnnnnnn nnnnnnnmmn nnnnnnnmmn nnnnnmmn nnnccgmmna agucaggaga
480
-continued

cugucugua ggccaaauug ccaec

<210> SEQ ID NO: 220
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium tumefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 220

cucuaugug gaaagcgacg guungccuac aacgcnnnnn nnnnnngaa nnnnnnnnn

nnggccgaag ggapnunnaa angggaacna uggugcgge gannnnnnuc uunnnnnnuc

ngccuauggc uggcgcgc ccccaacugu aangggaauu nnnnnnnnn nnnnnnnnn

nngnnnnnn nnnnnnnnn nnnnnnnngu uguccaucc ccggcggucu gaagggcuca

uunnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuguuuu

uunnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn

nngngcggaa gcggagggc nacgganmn nnnnnnnnn nnnnnnnnn

nngnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn

cugugcuca aaugggaaac caugc

<210> SEQ ID NO: 221
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium tumefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 221

cggauaauc ggcggugauc guuuccuucc gggnnnnnnnn nnnnnngcun nnnnnnnnn

nnunuugga aggggnaaa angggaacna gguuagggan nnnnnnnnca aannnnnnnn

ngccuaaucc gcggcgcgc ccccacacugu gagcgggnm nnnnnnnnn nnnnnnnnn

nnngnnnnnn nngggcggua aacggaaaucc ccaucggcaan nnnnnnnnn nnnnnnnnn

nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuucnnn

nnnnnnnn nnnnnnnnn nnnngccucc aucaannn nnnnnnnnn nnnnnnnnn

nnnnnnnn nnggggagggc aacggcgaggaga gguuacucn nnnnnnnnnn

nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncgunakan acggagga

cugccucaua ggggauacuc caugc

<210> SEQ ID NO: 222
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium tumefaciens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 222

gaccaugguu agccauccug guuncugcgg a cnmnnnnnn nnnnnngaa nnnnnnnnn

-continued

```
nnnnguucg gacgnaag cngggsuuc cngcugucg cngguccuc uccugguuggc ucnmncnmm c120
ngcugaaccu cngcgucgc cngcugucg cngcugucg cncggcngnc gncngcnmm cncncngnc 160
nnnnnnnnn nncnnnnnnn nnnncnmcnc gncngcncnc gncngcncnc gncngcncnc 240
nnnnnnnnn nncnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn nnnnnnnnn 300
nnnnnnnnn nncnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn nnnnnnnnn 360
nnnnnnnnnn cngcncncnc nnnncnmcnc gncngcncnc nnnnnnnnn nnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncnmcnc ngcngcncnc nnnnnnnnn 480
cccccgucu ucgucgcnc cnc 505
```

<210> SEQ ID NO 223
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium tumefaciens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 223

```
cccauagcuuc ucucgcuc uccucccuc uc uccnnnnnnn nnnnnnnnn nnnnnnnnnn 60
nnnnngug cncgncnnnc cngggsuuc cngggggg uccnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnncnmcnc uccggcnggc cncgucgcnc gncngcncnc gncngcncnc 160
nnnnnnnnn nnnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn nnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn 300
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn 360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncnmcnc gncngcncnc nnnnnnnnn 480
cccccgccgc cngcncncnc cnc 505
```

<210> SEQ ID NO 224
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium tumefaciens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 224

```
cuaaggggaa ggcggcgcng gnnncuucu uc cngcnnnnnn nnnnnngcnc nnnnnnnnn 60
nnnnnggugc ggcgccguc cngcncncnc gc uccnnnnnnn nnnnnnnnn nnnnnnnnn 120
ngcugcuaucc gcucgcucgc cngcncncnc gc uccnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn cngcncncnc gc uccnnnnnnn nnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cngcncncnc gc uccnnnnnnn 300
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cngcncncnc gc uccnnnnnnn 360
nnnnnnngg cngcncncnc gc uccnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cngcncncnc gc uccnnnnnnn 480
cccccgccguu caggaaaaa cgguc 505
```
-continued

<210> SEQ ID NO 225
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 225

auucaaucgu ugggaacag gunaeguua aguacnnnnnn nnnnaacuga uannnnnnnn
60
nnngacauua uguuunnnaa anggaaunc cgyucnnnnnn nnnnnnnnn nnnnnnnnn
120
nnnnnaauc gcgaeguccc ongcaacugu canuugennnn nnnnnnnnn nnnnnnnnn
180
nnnnnnnnn nnnnnnnnn mnnnnugag uguuacgau uuunnnnnnn nnnnnnnnn
240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn uacugacng
300
nnnnnnnnn nnnnnnnnn nnnnnuuca uuunnnnnnn nnnnnnnnn nnnnnnnnn
360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnngccuacng agccaggaga
420
ccguccggu cuuacaacac uguuu
505

<210> SEQ ID NO 226
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 226

uaguguuugu gacgggaacag gumgecnnnn nnnnnnnnn nnnnnccaaag cnnnnnnnnn
60
nnnnnnnnn gcguunnanaa anggaaunc uggucnnnn nnnnnnnnn nnnnnnnnn
120
nnnnnaauc gcgaeguccc cgcacaacugu ganugucunn nnnnnnnnn nnnnnnnnn
180
nnnnnnnnn nnnnnnnnn nnuuucacac uacugacnng nnnnnnnnn nnnnnnnnn
240
nnnnnnnnn nnnnnnnnn nnuuucacuc uunnnnnnnnn nnnnnnnnn nnnnnnnnn
300
ngagaaguc auugggaaac nguucuauag ugguaannnn nnnnnnnnn nnnnnnnnn
360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
420
ccguccguau uaguaacagaau uugcu
505

<210> SEQ ID NO 227
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 227

uagacacgcu caagcauuaa gungguuca annnnnnnnn nnnnaccauc cgnnnnnnnn
60
nnnnnnnnguacuuaacanggaauczgguganmnnnnnnnn nnnnnnnnn nnnnnnnnn
120
nnnaagucc agccagcugc cggcucagug auauaggcnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnnnnnnnnnn 480
cugccuauu guauugcaccg gcaa 505

<210> SEQ ID NO 228
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: ploc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 228
auugauauau ggcgagaggg guacgcuua cacnnnnnnnn nnnnnnnguuu nnnnnnnnnn 60
nnnnnnnnnn gucggguuana angggagngu cggangnnn nnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn gucggguu unncucucugc uauaugnnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnngag agccucagcag cagnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnncnnu cccagcunnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 480
cugccuauu guauugcaccg gcaa 505

<210> SEQ ID NO 229
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: ploc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 229
cggugaugaa uacngcagau cgcungnnnn nnnnnngacu gnnnnnnnnnnnnnn 60
nnnnnacagoc ggguanuaaa angggacgnc ccggcnnnnnn nnnnnnnnn nnnnnnnnnn 120
nnnnnacagoc gggugacgc cggcucagug auauaggcnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn 480
cugccuauu guauugcaccg gauuu 505
<210> SEQ ID NO 230
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bradyrhizobium japonicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 230
cguuaaucag aguucgcag guacucgcag uucnnnnnnn nnnnnnncau uunnnnnnnn 60
nngnagcg gacnnnaaag angggaangc cggugonnnn nnnnnnnnn nnnnnnnnn 120
nnnnaaacgc gguccgucgc cccgacacgu gancgcgnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng uccgugaugc 240
cnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnggcu ugggaggaac nccgaccacgc acgcagugnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nncaccgmcna acgccgaga 480
ccgcgccgca csuuaauug guca 505

<210> SEQ ID NO 231
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bradyrhizobium japonicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 231
caaaggggg ccocgcgugu guuaccguuc ucnnnnnnnn nnnnnncuaa uunnnnnnnnn 60
nnnnnnngac aggccgnaag angggaangc cguuaccggg eoccacugge aanggauugg 120
guccaaauu gcaagccgcg cccgacaccg gggcggagnn nnnnnnnnn nnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn nccgacccga gnnnnnnnn nnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cccacuacgc 300
cnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnngga ugggaggaac nccgaccacgc acgcagugnn nnnnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nncaccgmcna acgccgaga 480
cgccgacacgc ggacgacuuu uggac 505

<210> SEQ ID NO 232
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bradyrhizobium japonicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 232
gggcaccag gcccgggaug guuaccguca gguggcccmn nnnnnnnnn nnnnnnnnnnnn 60
nngnogccc aggccacuau cccgacaung gggaunggc ggaccencagug uccgmmnggc 120
gccacaacc cccgaccgcc cccgacacugu aangggunnn nnnnnnunnn nnnnnnnnn 180
-continued

```
nnnnnnnnn nnnnnnnnn nnnnnnnnnag gggcuccgaa cccnnnnnnn nnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnng ccacugggcc 300
nnnnnnnnnn nnnnnnnnn nnnnnnnng caannnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnnngu cggcgaaggc ccngccggac ccacugggnm nnnnnnnnn nnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nncocgngn agccaggaga 480
ccggcgugc auguuggag gcoca
505

<210> SEQ ID NO 233
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bradyrhizobium japonicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 233
 aaucuguau gcggcgcagg guuuccccc c nnnnnnnnn nnnnnnnnag nnnnnnnnn 60
 nnnnnnnnn ggaugnnnaa angggaagng cggugcgggg nnnnnnnnug uunnnnnnnn 120
 ccceaaagc gcggcgcagg ccccaacugu aangggnnnn nnnnnnnnn nnnnnnnnn 180
 nnnnnnnnn nnnnnnnn nnnnnnnncu ccuuuggccag aannnnnnnn nnnnnnnnn 240
 nnnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnnn nnnnnnnnnng ccacugggnn 300
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
ccggcgugc gcggguggca caacgc
505

<210> SEQ ID NO 234
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Bradyrhizobium japonicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 234
 ugguaguau gcggcggagc guuuccccc nnnnnnnnn nnnnnnnnag nnnnnnnnn 60
 nnnnnnnnnn agaunnnnn nnnnnnnng ugggcggaga uuggcuccac gcggcggaguug 120
 ucccaacggc cggcgcagg ccccaacugu aangggnnnn nnnnnnnnn nnnnnnnnn 180
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cccagcgggnn 300
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cccagcgggnn 360
 nnnnnnnnuc cggcgaaggc ccngccggac ccacugggnm nnnnnnnnn nnnnnnnnn 420
 nnnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
ccggcgugc gcggguggca caacgc
505

<210> SEQ ID NO 235
<211> LENGTH: 505
```
<210> SEQ ID NO: 236
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Brucella melitensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 236

agugacaaac acugacagc gumuucccg cgrnnnnnnnn nnnnaacga uccnnnnnnnn 60
nnnncgca acuangacnnn angggacnnn cggaaacc gcnnnnnnnn nnnnngg 120
cggacacg caccnnnnn nnnnagcng cagcnnnnn nnnnngg 180
nnnnnnnn nnnnngcnnn nnnnagcn nnncnnnnnn nnnnngg 240
nnnnnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 300
cnnncnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 360
nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 420
nnnnnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 480
cnnncnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 505

cgcnnnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 60
gacggcacg gacgcnacgc gcnnnnnnnn nnnnngcnnn nnnnngcnnn 120
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 180
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 240
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 300
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 360
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 420
gacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 480
cgcnnnnnn nnnnngcnnn nnnnngcnnn nnnnngcnnn 505

<210> SEQ ID NO: 237
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Brucella melitensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 237

gcgacaaac cgccacacgc gumuucccg ccgnnnnnnn nnnnngcnnn nnnnngcnnn 60
nnnncggc ggacggcacg gcnnnnnnnn nnnnngcnnn nnnnngcnnn 120
gcgacacgc ggacgcgacgc gcnnnnnnnn nnnnngcnnn nnnnngcnnn 180
gcgacacgc ggacgcgacgc gcnnnnnnnn nnnnngcnnn nnnnngcnnn 240
-continued

<210> SEQ ID NO 230
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Brucella melitensis
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 230

uaacaaaucc uguuguacg guucuuucg auuunnnnnn nnnnnngacn nnnnnnnnn
nnnagacgc gagcunnaag aeggaaauc gcyugcgcw gcocnmaaug gunnngggg
120
ggcaauggc gagcugccgc cgecaacugu aangegcgnn nnnnnnnnn nnnnnnnnn
nnnngacgc nagcgeocce caunnnnnnn nnnnnnnnn
240
nnnngacgc nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnng ccacucgcnn
300
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn
nnnnnnnnn
360
nnnngacgc nnggugagc nnguugcnn nnnnnnnnn nnnnnnnnn
nnnnnnnnn
420
nnnngacgc nnnnnnnn nnnnnnnn nnnnnnnn nnnnncgugng agccaggaga
480
cgcucgucu uacgagucg auugu
505

<210> SEQ ID NO 239
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Caulobacter crescentus
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 239

gccuggucgc cuguguucgc guncuucgcg aegnnnnnnn nnnnnnuucg nnnnnnnnn
nnnnngcgc gagcunnaag aeggaaacu cggugnegg nnnnnncug saannnnnn
120
cacaaucc gcgaccgcgc cgecaacugu aangegcgnn nnnnnnnnn nnnnnnnnn
nnnnngacgcганgegcgnn
180
nnnnnnnn nnnnnnnn nnnnnncgac gcguugccgu uucgunnnn nnnnnnnn
240
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnng uacugcgc
300
gccugacgc nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn uuuunnnnn
360
cggpgaucc ucggaagac cagggcagg cggagacnn nnnnnnnnn nnnnnnnnn
420
nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnnccugng agccaggaga
480
cgcucgucu cagauaucg cuooc
505

<210> SEQ ID NO 240
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Caulobacter crescentus
-continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 240

uacgcucag uacgcucag gurnucucun nnnnnnnnn nnnnnngaaas nnnnnnnnn
nnnnnnnn 60
nunnnnnnnga ggaugnnaas agggasacng agguwnnnn nnnnnnnnn nnnnnnnnn
120
nunngagac uccgcugccc ccgcacucugu acacgcgnmn nnnnnnnnn nnnnnnnnn
180
nnnnnnnn nnnnnnnnn nnnnccagac ucucgcucac aunnnnnnnn nnnnnnnnn
240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuggcgc
300
nnnnnnnn nnnnnnnnn nnnnnnnnnn ccacucuacaa aannnnnnnn nnnnnnnnn
360
nnnnnnnggc cuggagacgg ngacgccecca ggaacauuga cmmnnnnnnn nnnnnnnnn
420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng acacaggaaga
490
cucgcucgc gcgcuggcuc uucgc
505

<210> SEQ ID NO 241
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Chlorobium tepidum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 241

aucucuaua cgsausguug gurngcuccgc caugnnnnnn nnnnnngaaas nnnnnnnnn
nnnnnnnn 60
nnnncauaac ggcuwnnnaa angggaaunc oggugannn nnnnnnnnn nnnnnnnnn
120
nunngagac gcgcacucag gcgcucucgu aacuuuccnn nnnnnnnnn nnnnnnnnn
180
nnnnnggcuc gcgcacacag cuucucgu gaaucgcgc caacunnnnn nnnnnnnnn
240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuggcgc
300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn
360
nnnnnggac gcggccgaga uccnnnnnnn nnnnnnnnn nnnnnnnnn
420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnggaana agucaagaag
490
ccucucuaua auuuuuuggc uucgg
505

<210> SEQ ID NO 242
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Chlorobium tepidum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-462
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 242

guacucucuc gcacucagac gcacucgguu nnnnnnnnn nnnnnnnnn nnnnnnnnn
nnnnnnnn 60
nunnnnnnngc ggg ngaanaac angggaaenu acacggunnn nnnnnnnnn nnnnnnnnn
120
nunngauuc uacucucag ccgcacucgu acaacgcgnun nnnnnnnnn ccgcucgcaac
180
auucgcucgc ccacgcaucgc uucucugcgu uacucuucag
240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn uacucuucag
300
<210> SEQ ID NO 248
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 248

aguguaaua aasuaauuag gnuuuaaaau guuacmnnnn nnnnnnauuu nnnuuuuuunnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn...
-continued

guauauagc ucuaugugc guargaaug unnnnnnnnn nnnnnnaugu nnnnnnnnn 60
nnnnnnnaca uuucuganaa gaaagaaunu cggucnnnn nnnnnnnnn nnnnnnnnn 120
nnnggaucgc gaacugucgcc ccccaacugu aanggunnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nn unnnnnnnu nnnnggacaa gacugcagau unnnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cacugugac 300
unnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnggcu augggaacgu uucaugauug gaaagauuun nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
cuagccaaac aacgcaagagu caacu 505

<210> SEQ ID NO 256
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Mesorhizobium loti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 256
cuauauacoc gaaagccguc guunncnnnn nnnnnnnnn nnnnnnguuu unnnnnnnnn 60
nnnnnnnn ggacgccnaag angggagau ungacgcgc gaunnnnnuu ucnunnnnnu 120
gcogauguc gcagccgucgc ccccaacugu guunncnnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
cuagccgac aagccaaac ugcua 505

<210> SEQ ID NO 257
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Mesorhizobium loti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 257
gcucuaaauu ccucugcagc gcunnnnnnn nnnnnngcuaa cunnnnnnnn 60
nnnnngcgc gcucuauaac ungacggagc guunnnnnuu cuggnnnnnnu 120
ucucuaaauggc gcucucgucg ccacgacagu unnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnagac gcacgacacgc agnnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
ccugcgcuc ggcacaaag aaucc
<210> SEQ ID NO 258
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Mesorhizobium loti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 258
auagaucaug gacucucag gugncgcguu cgunnmmnmm nnnnnngacg nnnnnnnnnn 60
nnnnagggg cggagnnmaa ungagagnc cggunmmnmn nnnnnnnnmm nnnnnnnnnn 120
nnnnacaugg gcggcgcucc ccgaacagg gcgunmmnmn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnm nnnnnmmnmnn mmnnnmncga aacuugggcn nnnnnnnnnn 240
nnnnnnnnm nnnnnmmnnmnn nnnnnmmnnmnn nnnnnmmnmnn nnnnnmmmc ccaugggcn 300
nnnnnnnnm nnnnnmmnnmnn nnnnnmmmmmm nnnnnmmnmnn mncucncag agcnggaagaa 480
cacocacga cagaaaaaag ucgac
<210> SEQ ID NO 259
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Mesorhizobium loti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 259
gugauuguc gcaugauugc guuncuccgg gcggcinmmn nnnnnnmnmn nnnnnnnnmnn 60
ngcgcgacg gcgnnnnmaa angcagagnc cggunmmnmn nnnnnnnnmn nnnnnnnnnn 120
nnnnngaggg gcgcgcgcucc cgcacacgu uangccgngmm nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnm nnnnnmmnmnnm nnnnnmmnnm nnnnnmmmmmm ccaacccau uggunnnnnmn nnnnnnnnnn 240
nnnnnnnnm nnnnnmmnmnnm nnnnnmmmmnn nnnnnmmmmnn nnnnnmmnmng uacugagcc 300
nnnnnnnmn nnnnnnmnnm nnnnnmmggaa cngnnnnmmn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnngcc ugggagacag mgggcagag guwuagcmnm nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnm nnnnnnnmnnm nnnnnmmnmnm nnnnnmmnmnm nnnnncnnng agcaggaga 480
cucogcuca cgcacacggu ccaacg
<210> SEQ ID NO 260
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Mesorhizobium loti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 260
aagucgcgg ccacucgccg gugncccgcn nnnnnnnnn nnnnnmocga amnnnnnnnn 60
-continued

gaaaccagcc agecgcgcgc cccaaacgga aagngenam nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn agacgcnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn gcacugcug 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnngc auggaagcn nnggcgccg aagcgcagc cgcacuenga nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nngcgcgaga 480
cgcgcgcgc gcacgcgcgcaug 505

<210> SEQ ID NO: 266
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Pseudomonas aeruginosa
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 266

gcuauaugg gggugcugcg guanacgccc cccuuugcg nnnnnnnuaag nnnnnncggg 60
gccacaagc ggcccgaaag aacgagaacg cccuuugcg gcacuunnnu cggcncaggg 120
ggccacagc gcggcgcgcgc cccauagcg agacgacucu gcnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnngc gcacugcug 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn cgcggagggc cggcgccggc cggcgccggc cggcgccggc cgcgcgcgc 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nngcgcgaga 480
cgcgcgcgc gcacgcgcgcaug 505

<210> SEQ ID NO: 267
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Pseudomonas aeruginosa
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 267

cuauuauugg ggggugcgcgc cccuuugcg gcncnnnnnn nnnnnncggc cccuuuuu 60
nnnnncggc nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cggcncngcu 120
uggcuauugg gcggcgcgcgc cccauagcg agacgacucu gcnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnngc gcacugcug 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nngcgcgaga 480
cgcgcgcgc gcacgcgcgcaug 505
cgaugccuug cggguuqag guuuccucgc cguuuuunnn nnnnnngoga nnnnnnnnnn 60
nnnnnccgog gggcunnaag angggaacng cggugcnnnn nnnnnnnnn nnnnnnnnnnn 120
nnnnnnauggc ggcgcugcc cccgacacgu gacacggnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnccau cguuccccaa unnnnnnnnn nnnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cccacucggnn 300
nnnnnnnnn nnnnnnnnn nnnnnnnng nnnnnnnnn nnnnnnnnnnn 360
nnnnnnnnn ggggagagc nnnngggaacc ggggagaagc cccagannnn nnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncguung agcgagaga 480
cccacucug ccacacuucg cggc 505

<210> SEQ ID NO 269
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Pseudomonas putida
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 269

gcuacuacacu cggcxggcgcc gcnuucnnn nnnnnnnnn nnnnnnacca cnunnnnnnnn 60
nnnnnnnnn gnacuunnnn cunngggasnc ccccccgcccc ugnnnnnnca auunnnnnnca 120
ggcnnncnc ggcacacugc ccacacucugu 3cgnngnnnn nnnnnnnn nnnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnngcag cccacucuau cgaunnnnnn nnnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnng cccacuggen 300
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn nnnnnnnnnn gcucacnc ucagagaga 480
ccacucgcgc uacacucacc aacog 505

<210> SEQ ID NO 270
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Pseudomonas putida
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 270

cgacacucgc ccauucacu gunicccucuc gcnnnnnnnn nnnnnccgoc cnnnnnnnnn 60
nnnnnnncgc gggccmuuacu cncggccacnc cggccuucuc ugccnnnnnuug ccnnnnnnnca 120
cagacaaguc ccgucgccgc cccacccggu aangcgcggn nnnnnnnnn nnnnnnnnn 180
<210> SEQ ID NO 271
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Pseudomonas putida
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 271

uccccuggc ucgccuggac gusnccccnn nnnnnnnnn nnnnuuuucg nnnnnnnn 240
nnnnnnng gggugnnaaa cngggaanac cggyguguce caggccuuc aegnaagggcc 300
gacacagggc uggccugcc cccgaacggu angcagaggn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn gsaegcgcuc unnnnnnnnn nnnnnnnnn 420
nnnnncccc nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccacugccuc 480
cacggccgga guuacuggac accccc 505

<210> SEQ ID NO 272
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Pseudomonas putida
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 272
cgaugcuguu cccacuucug guuaccugc ccccccccc nnnnnnnnn nnnnnncugn nnnnnnnnn 60
nnnnnggcc aacunnnacg ccggacacng cggcuancnn nnnnnnnnn nnnnnnnnn 120
nnnnnnacgcc gggccucgcc cccgaacggu angcagaggn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cccacugccuc 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccccucuacnc 300
nnnnncccc nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccacugccuc 360
nnnnnnacgcc gggccacgac gggaccucgg cccacgcca aucgacacn cggccuacc 420
nnnnncccc nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cccacuucug 480
cacggccgga guuacuggac accccc 505

<210> SEQ ID NO 273
<211> LENGTH: 505
<210> SEQ ID NO 274
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Sinorhizobium meliloti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-468
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 274

cuagaagag gaccucucaag gugwcgcocu cnunnnnmm mnnnnngaag nnnnmmnnnn 60
nnungaaggg gcaacgnauu ugggaagnc ccgucannnn mnnnnnnnn nnnnmmnnnn 120
nnnnnaacc ggcgcgcgcc cccacaaggu gnnnggagen mnnnnnnnn nnnnmmnnnn 180
nnnnnnnnn mnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
nnnnnnnmm mnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnngu gccgacgccg cnggcgggc gnncgmmmm mnnnnnnnn nnnnnnnnn 360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
ccgacgacac agacacgacaa gaccg 505

<210> SEQ ID NO 275
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Sinorhizobium meliloti
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-468
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 275

uggcaauaug cggcgccag gugwcccgcn mnnnnnnnn nnnnnngaaau unnnnnnnnn 60
nnnnnnngc gggggnauu cgggaagnc ccgwcnnnnn mnnnnnnnn nnnnnnnnn 120
nnnnagcuuc ggcgcgcgac cccacgcucu gaaagggnnnn mnnnnnnnn nnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnngacg uucwgcacaa aaaggccucu gaaucuuc 240
agacuuanu nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuauua nnnnnnnnn nnnnnnnnunu ng gacuacuau 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnuaau ugcacuagcc uggccgagaa ccagcuccnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnuuuu nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnuuccuauc agucuagaga 480
cgcuccgc uggcugaccc gcgcuc 505

<210> SEQ ID NO 276
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Sinorhizobium mellotii
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 276

uacaauggc uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu nnnnnnnnn 60
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 120
cgcuccgc gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 180
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 240
aacuacgc uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 300
gacuacgc uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 360
gacuacgc uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 420
gacuacgc uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 480
cgcuccgc uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 505

<210> SEQ ID NO 277
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Sinorhizobium mellotii
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-468
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 277

cacauuacu ggacuacgc guguccuccu accnnnnnn nnnnnnguua nnnnnnnnn 60
nnnguugaggg gacuuuuuaa uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 120
gacuacgc uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 180
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 240
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 300
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 360
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 420
nnnnngaggg gacuuuuuaa uggcuaac gcucuaccg uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 480
cgcuccgc uggcuaac gcucuaccg uuuuuuccu nnnnngaggg uuuuuuuuuu 505
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 278

gcuaaccgacuccgagug unncuccgcc unnnnnnnnn nncagcugaa gacnnnnnnn 60
nnnnnnnn gc gguungnnn naa angccaacna ogguaggac gacnnnnnau cunnnnnnng 120
gguuaaccom guggcgcgcc cccgccacugu gacngggnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnccag caaguccaa gacnnnnnn nnnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnmg ccacgccccn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnnng cuagauagggc nggacacag ngacagcnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncgennn aegccaggag 480
cocgcaacu cccuggggca caacgc 505

<210> SEQ ID NO 279
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 279

uagcgucgcc cggcacaacg guuncgccccc gcccnnnnnn nnnnnnngcca nnnnnnnnnn 60
nnnggcggga ugcgccaagu aanggaacnc cggcgnnnnn nnnnnnnnn nnnnnnnnn 120
nnngnaucc ggcgcccncc ccccgccggg gacngggnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnncnca cccgccacau nnnnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnc gcacgcccc 300
cgnnnnnnn nnnnnnnnn nnnnnncnca uacncnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnngccgc ccgagcccg ccacgggcaac uagggcczu ccccgccacag gggggnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnscnng aegcgaaga 480
cocgcaacau gcccgegegc gcacgc 505

<210> SEQ ID NO 280
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 280

uagcgcagug ccgacgcug guuncgcgcc cucgcguccen nnnnnnacuac nnnnnnnngu 60
cucggcgc cgacgennaag aanggaacnc cggcgnnnnn nnnnnnnnn nnnnnnnnn 120
nnngnaucc ggcacgcucg ccggcgcggg guungggnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnngcagcgcgc 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnn 300
ccagmnnnn nnnnmmnn mnnnnnaug agnnnnnnnn nnnnnnnnn nnnnnmmnn 360
nnnnuugggc ccgggaecn mngcagcgcg guaggugcnn cnccgsccacu cguucccncg 420
uguagcgcgn nnnnmmnn nnnnnnnnn nnnnnnnnn nnnnnnnncng aguuccaga 480
ccugcgcagc cgcggcguag cguag 505

<210> SEQ ID NO 281
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEN: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 281

gcagcgcguu guacucagcg guunncaucgn nnnnnnnnn nnnnnncnogn nnnnnnnnnn 60
nnnnnnnnccg cnccgguanc cnccggcuunnn nnnnnnnnn nnnnnnnnnn 120
nnnngcngauuc gcggcgcucgu cgcggcagcu gcnnnggnnn nnnnnnnnnnn 180
nnnnnnnnn nnnnnnnnn mnngcagc gcnnnggunnn nnnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 300
nnnnnnnnnn nnnnnnnnn nnnnnnncgc cmmnnnnnnnn nnnnnnnnnnn 360
nnnnnnnnngc gcggcgcagc cggcgcagcn nnnnnnnnnnn nnnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnngnngc aguuccaga 480
cgcgcgcucgc gcggcgcguag ccucg 505

<210> SEQ ID NO 282
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEN: misc_feature
<222> LOCATION: 23-468
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 282

uaugcccaug cgcgcuucgc cmmnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnnn nnnnnngcag gcgggcnncn nnnnnnnnn nnnnnnnnnn 120
nnnngcngauuc gcggcgcucgu cgcggcagcu gcnnnggunnn nnnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 360
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnng ccacucgcgc 480
cgcgcgcgc gcggcgcguag ccgcgc 505

<210> SEQ ID NO 283
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Streptomyces coelicolor
<220> FEATURE:
<221> NAME/KEN: misc_feature
<220> SEQUENCE: 284

acuacuguc ocacugcug guanuunnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnn nnnnnuanu coccgcacuc cggugunnn nnnnnnnnn nnnnnnnnn 120
nnnnguaguc gggcagggc uccgacgucu ganaucgguu gnunnunnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnaacg ucggacgcuq gcgguaagcg gcanunnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng ccacuggauc 300
gnnnnnnnn nnnnnnnnn nnnnnncuuc gnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnogggc cggcaggcg gcggcaggca gcggcaggcu nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnunna agccaggaga 480
cggcaggcg ccagccuggc aucua 505

<220> SEQUENCE: 285

ccguacgcu cccugucog gumcunnn nnnnnnnnn nnnnnngugn nnnnnnnnn 60
nnnnnnnn nnnnnnnnn cggascasuc cggugcnnn nnnnnnnnn nnnnnnnnn 120
nnnnguacuc agacgugac gggcagaggu aangggan nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnncaacgg uccgacgguu ugguaagcgn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn cmmnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnng ggggaguc gcggcunnc ucggcagcu gcggcuucgn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnccgma agccggaga 480
cggcagccgc ccagccuggc aucgu 505

<220> SEQUENCE: 286

nuuugaguca cccuggugug ugncucuucgc augnnnnnn nnnnnnnauag nnnnnnnnn 60
nnnncucugcc gagaunnnau cggcaggcgc cagugannnn nnnnnnnnn nnnnnnnnn 120
nnnncuacuc gcugucugcc ccacacggcu aaagguunnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nngagagacgc gggcagmcunnn nnnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnunmc gggugucn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnaacg unnnnnnnnn nnnnnnnnn 360
nncnnngaa cccguuaac ucaguguc aaggcaguu uccgguuuauc cuauuguag 420
auuggauuua uouuuuunm nnnnnnnn nnnnnnnn nnnnnnnn nnnnnnnn agucggaga 480
cggogcuaa aggguuuuuu gaga 505

<210> SEQ ID NO 286
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Shewanella oneidensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 286

accuaugua uucuaauuag gucuuaaacc gcgcgannnn nnnnnnnnn nnnnnnnnn 60
uccacccaaa auunnnnauu angggauunc gcgctgccugm nnnnnnnccc gnunnnnnnnn 120
ncacgcacc gcacacugc ggcacacug gguuagnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cuagauucca 240
gauucgagn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ncucuaucnn 300
uagaucucag auucuaaagnuccagaccc uuwuuuunn nnnnnnnnn nnnnnnnnn 360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ncucuaacna agucggaga 420
cogcuuauu gguuuumug cuug 505

<210> SEQ ID NO 287
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Salmonella typhimurium
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 30-488
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 287

gcuccacau cuaccacacag guuuuccaccn nnnnnnnnn nnnnnnnnn nnnnnnnuuu 60
nnnnnnnnnng ggunnnnnnn angggaaeng gcggcaggnn nnnnnnnnn nnnnnnnnn 120
nnnnnaccc cccggccagcc cccgccgcugu gauunnnnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnnn nnnnnnnnn nnnnnnuga gcgcgguaa agannnnnn nnnnnnnnn 240
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn ccucuagcn 300
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
cogcuuucgc gguuauaaccc aaca 505

<210> SEQ ID NO 288
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Salmonella typhimurium
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 288
acggaagcau cpggggcgcg guucmcmcmn nnmmnnnnnn nnmmnmrgg nnnnnnnnn
60
nnmmnnmnn nagnmnanu angggaawnc cguvgmnmnn nnmmnmnnnn nnnmmnnnn
120
nnnnamaucc gggcggagac gcgcacgcau anngmmammm mmnmmmmnnn nnnmmmmn
180
nnmmnnnmm nnmmnmnnnn nnmmnmnmagg uggagagagc ggoaagcmaan nnnmmnmnnn
240
nnmmnnnnnn nnmmnnmmnn nnmmnmnnnmm mmnmmnmnmng acacnmcmnn
300
nnmmnnnnnn nnmmnmncmcmnmnn nnmmnmnnnn nnnmmnmnnn
360
nnmmnmng gggcggagcm naucamucuco gcguaccacgc ccacggacg ccmmmcmcmnnn
420
nnmmnnnnnn nnmmnmnmmnn nnmmnmnmmnn mmnmmnmncmcagc agcccggaag
480
cgcgcgcgcc agccgcgcau cgggu
505

<210> SEQ ID NO 289
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Thermotoga maritima
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 289
gacgccucuc uacccggagcg gumcacccmm nnmmnmnnnn nnnmmnmucg nnnmmnnnn
60
nnmmnmnmng ggucmcmcmn gngggaagnn cmcmnmnnnm nnmmnmnmnn nnmmnmnmnn
120
nnmmnmamaucc gggcgcggngcm gcgcacgcmnm gmcgggcmn mmnmmmmmmnn nnnmmmmn
180
nnmmnmnmnn nnmmnmnmnn nnmmnmngcm acacgcgac acnmnmnmnmnn nnnmmnmnmnn
240
nnmmnmnmnn nnmmnmnmnn nnmmnmnmcm gacmggcmnc uggmggcmnm mmnmmnmnmnn
300
nnmmnmnmncmc ugggaagcm nmcgccggag uggagagcmn mmnmmnmnmnn nnnmmnmnmnn
360
nnmmnmnmnn nnmmnmnmmnn nnmmnmnmnmn mmnmmnmnmncagc agccgggacag
420
cgcgcgcgcg guccagggga accac
505

<210> SEQ ID NO 290
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 290
uacgaauua aagccuaag guncoccmnn nnmmnmnmnn nnnmmnmnnu uggmcmnmn
60
nnmmnmnmng gguwnnnnn angggaagac gggmcmnnmm nnmmnmnmnn nnnmmnmnmnn
120
nnmmgaaucc gcgcgcmcmn ccgcucgcmn gacngcmnmnn nnmmnmnmnn nnnmmnmnmnn
180
nnmmnmnmnn nnmmnmnmnn nnmmnmnmnn mmnmmnmnmng ccacgcgucc
240
gcgcucacu ggcgcgcmm uacguaggc gsoagaagc agasacmcu uccguacgu
300
ugacucgcgg gggggaggc naccggcc uggagacnmn nnmmnmnmnmnn
360
auaguaugcg cuucaacgcg guungcuauc ugnnnnmmnnnn nnnnnngauu annnnmnnnnn 60
nnnnnngauu gycuyunnnnn anggaacncu cgguyunnnnn nnnnnnmmnnnn nnnnnnnnn 120
nnnngnangu ggaacagagc ggcagcgcggauu anaagagnnuu nnnnnnmmnnnn nnnnnnnnn 180
nnnnnmmnnn nnnnnnnn nnnnnnaac naggcuuaa unccnnnnnnnn nnnnnnnnn 240
nnnnnnmmnnn nnnnnnnn nnnnnnnn nnnnnnnnng acacugcag 300
nnnnnmmnnn nnnnnnnn nnnnnngga nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnngnauu guggaagcuc mnggacagu nguuaacag nnnnnnnnn nnnnnnnnn 420
nnnnnmmnnn nnnnnnnn nnnnnnnn nnnnnnnn nnnucunug agugcgaas 480
cucgcagcua cugcagcaca cacug 505

<210> SEQ ID NO 294
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Xanthomonas campestris
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 294
cuucaacgcg egcccccugag gugnccucgc gpnmmnnnnmm nnnnnnaauu nnnnnnnnn 60
nnnnccgggu gnnnnnnnnaa cnggacacuc cggugcgcgcu ugnmmngcag 120
agcagucgc ggaacagacgc cggcaacgcgu ggnncgaagnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnmmnnnnn nnnnnnnnn nnnnnngcu aagccagcaca cagnnnnnnn nnnnnnnnn 240
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnn nnnnnnnnn nnnnnnnng ccacugcugnn 300
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cccgcucnng 360
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnccugnng ccgcgcag 420
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cagccggaga 480
cagccugcug gauugacccg ggcac 505

<210> SEQ ID NO 295
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Xanthomonas citri
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 295
cuucaacgcg egcccccugag gugnccucgc gpnmmnnnnmm nnnnnnuguu nnnnnnnnn 60
nnnnccgggu gnnnnnnnnaa cnggacacuc cggugcgcgcu ugnmmngcag 120
cucgcacauc ggaacagacgc cggcaacgcgu ggnncgaagnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnmmnnnnn nnnnnnnnn nnnnnngcu aagccagcaca cagnnnnnnn nnnnnnnnn 240
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnn nnnnnnnnn nnnnmmng ccacugcugnn 300
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cccgcucnng 360
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cagccggaga 420
nnnnnnmmnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng cagccggaga 480
cggccgag ggauagcccg ggcac 505

<210> SEQ ID NO 296
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Yersinia pestis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 296
uuacuagc uacchuagc gucgccgcuc augcguuun nnnnnnaauu annnnnnnnn 60
naacacuua gauunnaaa angggauun ccggugunnn nnnnnnnnn nnnnnnnnn 120
nnnnnaauc ggcgugacn gcggcgcggg ugggganu nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng acacugucn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnac nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnng nnnnnnnnn 360
nnnnnnng ggggaguc nnnncugccg cuuuaauucc ggccauun uuauacacag 420
auuuuuacg ucuacacuuu gcggcguauuc cagagnnn nnnncunnnn agcgcgaaga 480
cucgcgguu uacgcguuca auuuu 505

<210> SEQ ID NO 297
<211> LENGTH: 506
<212> TYPE: RNA
<213> ORGANISM: Acinetobacter calcoaceticus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 30-470
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 297
cuuacacaa ucucaacaa guuaaagcnn nnnnnnnnn nnnnnnauuc nnnnnnnnn 60
nnnnnnngc uuuunnnnnn anggggaunc ugggganun nnnnnnnnn nnnnnnnnn 120
nnnnnaauc gcgcgcuucgc cgcgcgcgg ugcuuaagn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnnuu uacacauuu aaaaagucacauauacauuuan 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnccgcacucau 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
accgcuguuu uacacauuc caucuca 506

<210> SEQ ID NO 298
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Agrobacterium vitis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 298
cusaaaggg cagcguauuc guuuucugca agugunnnn nnnnnnccaa nnnnnnnnn 60
-continued

nnccgcgc ggauuunnnn angggaaaau cagugaggc gacconmaag uaaaaannnnng 120
ggccgaacc ggugucgcc cccguacgcu gaaacagguu amnuuuuunn nnnnunnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnnn 360
nnnnnng cgguaagggc nngcacaag cacagacnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnngcunng agccagaga 480
cucgcuaua agcauucgcgc aaagc 505

<210> SEQ ID NO 299
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Bacteroides fragilis
<220> FEATURE:
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 299
uuauuucug ucuccuaucg guuuucgaa uacnnnnnnn nnnnnucauu ccunnnnnnnn 60
nnnnncucc gguauuunnnn anggaaangc ggugucnnn nnnnnnnnn nnnnnnnnnn 120
nnnnncucc gguauuunnnn anggaaangc ggugucnnn nnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn cccugacag 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn cccugacag 300
accuuccug nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
uaccagcuca cccguacggc nnnncgcuauu ccnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnn cccgcuauu 480
cucgcuaua aucacaggc guuuc 505

<210> SEQ ID NO 300
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Bacillus megaterium
<220> FEATURE:
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 300
auccaaacgc aaacagauag guuunccnnn nnnnnnnnn nnnnnnaaga amnnnnnnnn 60
nnnnnnnnnn gguuuuauu angggaaaac gguuunnan nnnnnnnnn nnnnnnnnn 120
nnnnngagcc uccuuuguuuu cccguacgcu aangugnu nnnnnnnnn nnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 480
cucgcuaua uugacgguu ucuaau 505
<210> SEQ ID NO: 301
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Leishmania major
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 301

nnnnnnnn nnnnnuugg gugncgccnn nnnnnnnnn nnnnnucac nnnnnnnnn 60
nnnnnnnn ggguguannya ccgggaaac ccggagguca uguuccuua cuccaaagggc 120
ugacgaguc cgcugucgcc cggacacggu aangcgagnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnug aagcucaaa uanmmnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cccuugugcc 300
nnnnnnnn nnnnnnnnn nnnnnncca gnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 420
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnn 480
cggccgcga aaaccagau ccacaa 505

<210> SEQ ID NO: 302
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Propionibacterium fraudenreichii
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 302

uguguguguc uguugugugc guunuggugc cunnnnnnn nnnnnnnocac nnnnnnnnn 60
nnnnnggug cuguugcagaug gnnnacuuc cggugyunn nnnnnnnnn nnnnnnnnn 120
nnnnauucu gagaagucgccc ggacagcgggu canruuggnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnn nnnnnnnnn nnnnnnnaac gcaacacggcu aagnnnnnnn nnnnnnnnn 240
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 300
nnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 360
nnnnnnncgc cgggaagn nuguagugga ggaugucgg agaugucug caugnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnc nnnnnnnnn nnnnnnnnn 480
cggccgcga gcccacacaau cgguu 505

<210> SEQ ID NO: 303
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Rhodobacter capsulatus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-468
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 303

gccacucggg gogcgccggu guuucuguc nnnnnnnnn nnnnnnnncuaau nnnnnnnnnnn 60
nnnnnnncac aggccnaaag aggggaang ugaaggggaau ugcagcggcu uunngcccg 120
aaacccgacc gcagcgcgccc cgcgcgcgug gcagcgcgcgc gnnnnnnnnnn nnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnng cccgccgcgc nnnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnng cccgccgcgc nnnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 360
nnnnnnnnnng cgcgggaaggc nngggggggc cggugagggaga ccccccccccg ccannnnnnnnn 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncncnca agcgggaga 480
ccggccgccc ccgggaauucc gggg 505

<210> SEQ ID NO: 304
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Rhodobacter capsulatus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 304
gggcuacucc acgcgcggaug guuucccnn nnnnnnnnn nnnnnacugg acnnnnnnnn 60
nnnnnnnnng ggcannanau angaaccna cggugagggau acnnnnnnnn nnnnnnnng 120
gggccuaucgg gcggggggggc cgcgcgcgug gcagcgcgcgc nnnnnnnnn nnnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnncnca cccgccgcgc 240
nnnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnncnca cccgccgcgc 300
ccccggccgg gcgccggccgc cccgccggcgc cggccgcgcgc cccgccgcgc 360
gggagaagggc ggcgggaggc nngggggggc uggagcgaan nnnnnnnnn nnnnnnnnnnn 420
nnnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncncnca agcgggaga 480
ccggccgccc ccgggaauucc gcgg 505

<210> SEQ ID NO: 305
<211> LENGTH: 505
<212> TYPE: DNA
<213> ORGANISM: Rhodobacter capsulatus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 305
gggccacccc gcggcgcggaug guuucccgcg ccagcnnnnnn nnnnnncacn nnnnnnnnnnn 60
nngccgccc gcggcgggaa cggccgacng ggcgcgggcu uacnnnnnnng nnnnnnnng 120
gggcccaacgcc gcgggccccgg cggccgcgcgc gcgcgccgcgc cccgccgcgc 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn cagcgcgcgc cgcgcgcgcgc cgcgcgcgcgc 240
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnnncnca cccgccgcgc 300
ggcgggggg cggggggggg cggggggggg cggggggggg cggggggggg 360
gggccgccc gcgggagggc nngggggggc cgcgcgcgcgc cgcgcgcgcgc 420
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnncncnca agcgggaga 480
ccggccgccc ccgggaauucc cggcc 505
<210> SEQ ID NO: 306
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Rhodobacter sphaeroides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 22-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 306
uguuuugg cagggucag gngmccgcn mnmmnmmn nnnnmucnq nnnnnnnnn
nnnnnnnng egsnmmmmn nmmgcsagc nnnmmmmmm nnnnnnnnn
nnnnnnnn ccssccggscc cssccgscc ggcsscggc gcangmmmm mnnnmmmmn
nnnnnnnn ccsscggcss nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnmnnnc ccssccggc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnmnnnc ccssccggc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnmnnnc ccssccggc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
ccggccgac ccssccggc ccssccggc 505

<210> SEQ ID NO: 307
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Sorghum bicolor
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 307
uagccugcgc ccacuccuag gugmccguc ggnmmmmmm nnnmmmmmm nnnmmmmmm
nnnncggc ccssccggscc ccssccggscc ggcsscggc gcangmmmm mnnnmmmmn
nnnnnnnn ccssccggcss nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnnnnn ccssccggc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnnnnn ccssccggc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnmnnnc augggcagcc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
nnnnmnnnc augggcagcc nnnmmmmmm mnnmmmmmm mnnmmmmmm mnnmmmmmm
ccggccgac ccssccggc ccssccggc 480
ccggccgac ccssccggc ccssccggc 505

<210> SEQ ID NO: 308
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Streptomyces griseus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 308
uagccugac cgggcaucg guuncgcucu gucnmnmnm nnnmmgcac nnnnmmmn
nnnngcagc cgggcaucg guuncgcucu gucnmnmnm nnnmmgcac nnnnmmmn
nnnnncgc cgggcaucg guuncgcucu gucnmnmnm nnnmmgcac nnnnmmmn
nnnnmnnnc cgggcaucg guuncgcucu gucnmnmnm nnnmmgcac nnnnmmmn 180
<210> SEQ ID NO: 309
<211> LENGTH: 505
<212> TYPE: RNA
<213> ORGANISM: Stealth virus
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: 23-469
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 309
auucgucaucgguaggagcpgnnucucggccnnnnnnnnnnnnnnnnnagacnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 313

aataaatcga aacatcatt tcgtatastg gcggaatag ggcctgcgga gtttctacca 60
agctacgta aatagttcng actacgaaat taatgggttt tttata 105

cgtttttttat ataaagtacc tcataatact ttggaatat ggcocaaaaa gtttctaccc 60

gctgaacgta aatcggcgcgn aactagggga aagattttgg atctt 105
<210> SEQ ID NO 315
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 28-79
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 315

```
ttaatcagct tcacactct tcgtaaatt gcggatcagg gttctacag
```

60
gtannccgta aataccttna gctacgaaaa gatgcgaatt aatgt

105

<210> SEQ ID NO 316
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 316

```
attacatta aaaaaagac tcgtaaatt gcggatcagg ggtctcgaia gttctacaa
```

60
ggctgccgta aacagcgtgn actacgagtg ataccttgac ataga

105

<210> SEQ ID NO 317
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 317

```
agaatcata taagatgat tcgtaaatt gcggatcagg ggtctcgaia gttctacaa
```

60
gctcagctga aatggcctgn actacgtaaa cattttcttc gtttg

105

<210> SEQ ID NO 318
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 318

```
catgaaacca aaaaaagac tcgtaaatt gcggatcagg ggtctcgaia gttctacaa
```

60
gcgaacgta aatggtcgtn actatgagg aaagtgtagc ataaa

105

<210> SEQ ID NO 319
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 319

```	ttcaataa taaggaacac tcgtaaatt gcggatcagg ggtctcgaia gttctacog
```

60
---continued---

ggcanccgta aantgccgn actatgggtg agcaatggaa cggca 105

<210> SEQ ID NO 320
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 43-80
  <223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 320
catcctagaa aasagacctc ttgatatag tcagtaaatag ggtctgtgatt gtttctacct 60
agtaacgcta aaaaaactgna actcaaaaaa gtttggtata aattt 105

<210> SEQ ID NO 321
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 29-80
  <223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 321
tatataaaa actaaatattc tcgtatamca cccgtaaatag ggttccggaa gttctctacct 60
gtcgccacta aantagcgag actacggggtgt tttattgata atata 105

<210> SEQ ID NO 322
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 43-80
  <223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 322
gaaaaagttat aacatattac cccgtagtg ttgaaatag ggttctaaag gtttctctc 60
gactgcggta aantgtctgn actatgggtg tttataagta tttta 105

<210> SEQ ID NO 323
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 29-80
  <223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 323
aatgcggact atagtttaaac tcgtatatnt tcgtgaatat ggcncaggag gtttctacaa 60
ggaanctta aantttottn actataggtg attgttttgt atgca 105

<210> SEQ ID NO 324
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: 43-80
  <223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 324

tatatctta taataagata tcgtatatgc tcgaagatat ggnntgtggt gtttctacta 60
ggaggcogta aacpctctan actacgaata tataggtgat ttctta 105

<210> SEQ ID NO 325
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 325
taagttgatt aatattttaaac tcgttataaa tcggtaaatg ggntcggaga gtttctactc 60
gtacacgta aasatgcgag actacgaga gttgtactat aaatt 105

<210> SEQ ID NO 326
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 29-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 326
aaaaacggat atataacacta tcgtataang tttgtaataa ggnntcggagc gtttctacgg 60
gaancccta aaattcoggn tctatatggt aatgtgtat actat 105

<210> SEQ ID NO 327
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Fusobacterium nucleatum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 29-73
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 327
ataataatt tataaaaaat tcgtataang tttataatg ggntaaggt gtcttacgg 60
tttaatccata aaatnttaaca gtcagaaaa atgttttacct gttt 105

<210> SEQ ID NO 328
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Lactococcus lactis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 28-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 328
gtttataa gaaactattt atttattnn octaggtat ggnntcgggc gtttctacct 60
cgtancogta aatncgagng aacaataagga aattgattt tttag 105

<210> SEQ ID NO 329
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 329
aatccgctac aatatatag tcgtatasgt tcgtatatat ggnacggttc gtttctacca  60
gccaagcga aatgacagct gtgtgaatat attgtaaat ttaat  105

<210> SEQ ID NO: 330
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Listeria monocytogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 330
ataatataaa accgaataac ttgataataa gttgctattg tgggacgga gtttcctacot  60
gttacgcta aatccgctga cattgtataat aaggg  105

<210> SEQ ID NO: 331
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 331
caaatcattt ccatgctt tcgtatatcc tcgataat ggtttcggaa gttctacog  60
gttacgcta aatgatcgtg aactagaaag cagaagcagg ttcgg  105

<210> SEQ ID NO: 332
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 332
tgtgtaatt gaaatgaaat ggtgtataatt aagggattg gnncccaca gtttcctacca  60
gaccacgta aatggtctgg aactagctagt aatatattt gtatc  105

<210> SEQ ID NO: 333
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 333
cgcacatgg aaatgaaacc tcataataat ttagaataat ggnctcagaa gtttcctaccc  60
agcagctgta aatgctgctg aactagagggt aagattgatat atttc  105
<210> SEQ ID NO 334
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 334
aacctttata tattgttttt tcataatac gcgggatat ggncttgcaaa gtttctacg
60
gtttacgta aatgaacggn actatggaaa agcggaaaaat tagt
105

<210> SEQ ID NO 335
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 335
gttataaat ttaataaacc tcataatac taagaatat ggotttggaag gtttctacca
60
tgtggccttg aacgacatgn actatgagta acaacacaaat actag
105

<210> SEQ ID NO 336
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 336
cataaaaaatttatagac tcataatac taagaatat ggotttggaag gtttctacgc
60
ttgtgcctga aacgacacgn actatgagta acaacacaaat actag
105

<210> SEQ ID NO 337
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Streptococcus agalactiae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 29-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 337
catataata tatgatattc tatcttatng ctggaagnet ggnnc ttggt tgcggtccg
60
gacancgn aantgtoan acaataagta agcataaaat tagt
105

<210> SEQ ID NO 338
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Streptococcus pyogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 29-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 338
tgaatctaat aatgacatac tttatctng ctggtaatng ggnncgc ggcgtttcagc
60
gttctacaa
gacancccatt gatctan acaataagta acgttttagg cttgc 105

<210> SEQ ID NO 339
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Streptococcus pneumoniae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 29-79
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 339

aaaaattgga atcgtttttc ttggttatng tcgtaaatnt ggtncaacagac gtcttacaga 60
gtgtgncnggg aacacotna acaataagta acgtcagcagt gagat 105

<210> SEQ ID NO 340
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Thermoanaerobacter tengcongensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 340

aaaaattttaa taagaagacac tcataatact ccgagaatat ggntctcgga gtccttacctg 60
aacaacgta aatggtctcg nctatgagtg aacagtgacc taggg 105

<210> SEQ ID NO 341
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 341

aaaaatattag ctatatcgc tctgataacc tcaataaat ggnttctagg ggtcttacca 60
ggacanagta aasctctgmn attacaaaat ttgtttatga cattt 105

<210> SEQ ID NO 342
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 43-80
<223> OTHER INFORMATION: n = g, a, c or t/u

<400> SEQUENCE: 342

ataaaaaaat aasatctggct tctgataact ctaagatact gnnnttagag gtccttacca 60
gagacngag aantcttctgn attacgaaga aagcatttt ctttt 105

<210> SEQ ID NO 343
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Vibrio vulnificus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 50-80
<223> OTHER INFORMATION: n = g, a, c or t/u
<400> SEQUENCE: 343
gcctttgyc gatcaagct tcatataatc tcatgatat ggttgggag acctaacca 60
agagcttta aanccttgg attatgagt ctgagtctt atgct 105

<210> SEQ ID NO 344
<211> LENGTH: 229
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-201
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 344
agagaugguu gaggungcga aaacnnmaag naguacmca caguagagaga aaugnmnmag 60
aunnmmcg ugacnnmnnca cuguungggaa gggnngggauu ggccgacagg cagaucgaggg 120
ncuacauuccc nnnnnncgu gagcuauguu unnnngasan aacauagggc ugcacaca 180
cawnncnn cocaaannnn ncuacugag uggagacca uucacgu 228

<210> SEQ ID NO 345
<211> LENGTH: 228
<212> TYPE: DNA
<213> ORGANISM: Vibrio vulnificus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-203
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 345
agagagaua gaggungcga aaacnnmaag naguacmca cauugaaaaa gggnnngauu 60
gagannnnnc cguugagaaau uguunnggaaa gggnngggauu ggccgagacc ggaacaauun 120
ncuacauungu cguagagcga cguacuagauu unnnnnauu aacauacgaa ucuacuauag 180
cgnnnnnnnn ggunnnmmnn nnnnncguu uggagggcua uucacgc 228

<210> SEQ ID NO 346
<211> LENGTH: 228
<212> TYPE: DNA
<213> ORGANISM: Bacillus halodurans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-206
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 346
agagagagag gaggagagc guuuacuagag naguuaangc cuugnnnnnn nnnnngagaug 60
acacagga cnnnnnnnnn gocnccgaa ggnnnaccau ggcocagag cggagauun 120
aguacacgag cguucuacggu gggguugcau unnnngasan ahuacuacccc cuacacägcn 180
nnnnnnnnn gauunnnnnn nnnnncgcgu uggagaccau cuacggu 228

<210> SEQ ID NO 347
<211> LENGTH: 229
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-205
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 347

ggugagaua gggugucga aacucnaag neguauagcc uuuggagaaen agannnnnug  60
gaunnnnnnu cugugaanaa aggcnugaa ggggagcgg uggcgaagca aauuuuacn  120
ncacuccngu uuuuuuuucu ggcggugcug uunnggaaan auuggaaggc ugucaagaa  180
nnnnnnnnu caunnnnnnmn nnnnnnnucu uggagggcua ucuugu  228

<210> SEQ ID NO 349
<211> LENGTH: 228
<212> TYPE: DNA
<213> ORGANISM: Clostridium acetobutylicum
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-225
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 349

acccuugugua gggucgcuu uuacucnaag neguaucccg uuugnnnag uunnnnnnnng  60
gacnnnnnnu cuggucagaa acgcuccuaa ggggucggcuu uggcgaagca uuuggaanun  120
ngcannngu uuuuuuuucu ggcgguggcuu cunnncacan uuacuagugc ugucaucuua  180
uuacucuunnu cuuunnnnnnn uugggcaagcuu uugggcuu uuggacccua caannguu  228

<210> SEQ ID NO 349
<211> LENGTH: 228
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 6-208
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 349

aaaganguu gggcngcga gaaacncnaag ncaacuccua aauggganu uunnnnnnnnu  60
agnnnnnnag cugaacgcuu uuguunngaa ggggaaauuu ggcccagauu uuuggcuuaa  120
uccacuanng gcuaacugcug ggguggugcua annggaaan uuacaacac ugucaacnnn  180
nnnnnnnnnn uuuunnnnnnn mnnnnnnnug uggagccua ucaacuua  228

<210> SEQ ID NO 350
<211> LENGTH: 229
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-207
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 350

gaccaagaua ggggucgcuu uauuunnaag neguaucccuu auuggagcuu gacnnnnnnu  60
agunnnnnu unuacagcua uuancngga ggggaaauuu ggcccagagau auuuuauuun  120
ngggcnnmg uauuuuuucu ggggaaugc uunnnnnun unuggcanaa uggcacaucuu  180
nnnnnnnnnn uuuunnnnnnn mnnnnnnnug gggagcgcaccaacgu  229

<210> SEQ ID NO 351
<211> LENGTH: 228
<212> TYPE: DNA
<213> ORGANISM: Clostridium perfringens
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-206
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 351

aaccgaaga gggcngcgq ugnauuaauu nguannmuuc uuggcagggg uunnunnnnn

agcnnnnnn nnaugagaac aaagnugaaag gnnaaagaau cgccgaaacc uunuaaga

ggcuuuaau cuuiaaggu ggggguucau anmngaauan uauguaacuc uguacaaan

nnnnnnnnnn uaanunnnnn nnnnnnng uggugucua ucaugaa

<210> SEQ ID NO 352
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Escherichia coli
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-167
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 352

caggccgaga gaggcngcgq unuucccann ugguaacggu uugygmnag gannnnmmn

ccagnnnnnn cccuguaaa caacnnnnnn uggcuuucguu cggcggaggg uuggaagcng

caggccnccg uccacuucac gcccuaaggg ccccugngggu uuggacacga

aggccgcucg uggcggcguu ucgcagcggg uggagccuuu cugggaga

<210> SEQ ID NO 353
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Haemophilus influenzae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-205
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 353

uacacagaa gaggcngcgq uuauunaauu ugguaamuuc uuucagagmu gnnnnnnnn

auanennnn cggcagqaaa aasangaaa ggnnaauuu uggcggaaac aaaaaaann

nguucnmmu ugguguuggu uggcggcucg ucnngaang ggcggcacac uguacaagu

nnnnnnnnuu uggcuuunnn nnnnaacua uggacgcuu oggguuu

<210> SEQ ID NO 354
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Oceanobacillus iheyensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-205
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 354

guuuuggaga ugggngcgg agaaccnauc ugguaumuugg uccgcggammn aggynnnn

ugngnnnn ccuuguagaggg uuguunnnnaa uggcggacagc gugunnnn

auacauacauu uacacugguu gggcgcucua uunngaacaa auacagucu uguacnauag

nnnnnnnnnnn agannnnnn nnnnnccuaa uggagggcua ucggacug

nnnnnnnnnnn agannnnnn nnnnnccuaa uggagggcua ucggacug
ucgugguya gaggangcau acaacnnauu naguarnauug gacnnnnnnn naagagaug 60
aaacgauga uannnnnnu ugggunnggaa gqppnguuguu ugcggagaga cuaaauaann 120
ggucagacuu auaaauugcu guacuacuucu unnnagaauu aagagacac ugucaugac 180
nnnnnnnaaa aauaagnnnn nnnnnnnnca uggagaca cguacac 228

ucuaugugya gagganga vaa ucaacnnauu naguarnauu uuucugagmu gnnnnnng 60
auanannnn cgsagaggsa asaangnnnva gqppnguugac cggcgcgac uccugcuann 120
nguacnnnuu uugaauuggu ugugcguacu unnnagaang ganaogcuac ugucaaugun 180
nnnnnnuca uuuaaasnnn nnnnnnnca uggagcucu cuugguugg 228

auaauuugaa gagggcngcau canacnuag uguarnuug uuuagaauu anmmnnncau 60
uuggnnnnnnaa uccoacvga uauuunnnnaa ggnguugcga ugcggagac cuaauuaaun 120
ngacnnnuu uuuaauugaa ggacucuugug guuunagaag cuungcagaa ugucaauaau 180
nnnnnnnnnnauuuuan nnnnnnucca uggagugcau caacuugg 228

auaauugaua gaagggcguca guuuaauuuu naguarnacu uguunnnnca gaaagguauu 60
uccacuaccu suucnnnnac suuunnnnaa gqppnguugaa uggcgcgaa ucsaauaann 120
cacaamnnn uuauaucuauu gggagacguuu unncgauauc gaaacuacgac uggacacnn 180
<210> SEQ ID NO: 359
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-206
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 359

agauuuugau gagccgcau canaucnaug naguannaaac uuuagauaaau uugnnnucug 60
cuannnnca anuuanuug adgunnnaaa ggngngaugga ugccgaaauu gauuaaauu 120
nacnnnguu auggauuguu ggcaccuaua ugguuuagc cauunaaggu ugcuauaacu 180
nnnnnnnn nuaannnaua ugagagacau caacuugua 228

<210> SEQ ID NO: 360
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Staphylococcus epidermidis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 26-223
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 360

aaacagcuua gagggcgcgu auuaanuaua naguannnaaa uauannnnca gaacugcuau 60
gggacacug uggannnnca aauannnaaa gnnnnaaauguggacgauug auguaauucu 120
nccacuannau ggcgcaccau ugcacacacu uggucacauac cuuaccacacg uacnnn 180
nnnnnnnn nuaannnaua ugagagacau caacuugua 228

<210> SEQ ID NO: 361
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Shigella flexneri
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-167
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 361

cagccacagaa gaggggcgcgu unugcccccnn naguacgacgu guuugnnnag gannnnnnng 60
cagnnnnnu ccuguggcua caaannuuaa gggguggcac ccggcagcgc uuggagcng 120
cagccacacg uucacacacg ggccuaggc gncuacgcu ccccugnggu ugcacacaga 180
agccacuucu cggccgcgu uggccagcgg uggccacuu uggcnguac 228

<210> SEQ ID NO: 362
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Shewanella oneidensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-208
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 362

aggaacagaa gagggcgcgu uaacucuaua ngguannguc aacagcannn ggagnnnnca 60
-continued

cacaamnncuc cagcgagauu ugaunnnnag ggnagauuag gcgcagaugguuga uagauggunn 120
gguccguuau uguuaaauuc gguccgauug gnucuauumnn ccucuaaccgau uguuaccomnn 180
nnnnnnnnnnnnnugnnnng gugagaagcu uggugac 228

<210> SEQ ID NO 363
<211> LENGTH: 229
<212> TYPE: RNA
<213> ORGANISM: Shewanella oneidensis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-206
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 363
ccuuaaugua gaggncgogc ugcuccTable cannot be generated.
-continued

<400> SEQUENCE: 366

cgcaauaua gaggangcug ccagcnaun nngauauugg ccagcuuguu aaggagaagas 60
cucnnnbn nnaaanucg gcugnnaagaa ggnnauugg gggcgaaggg guggacuung 120
nnuunnnuga gcuccuccu gugguuaac nnnccaamn nnguuecaco ugcuauggga 180
nnnnnnnnc cccnnnnnn nnnnucacc auggacgcu uuauagca 228

<210> SEQ ID NO 367
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Vibrio cholerae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-206
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 367

ucuacgacaa gaggangcac ugnncccaag ccagacuuuu uuggouunnn cccnnccucca 60
acucnnnaun nnnnnnnnn nngaacaucca ggggagaugg uggggagggga aaauuacau 120
ngnngngcc uuuccuuaae gguuagcaggg nccnncuuac uccnuuacu ucuaucagn 180
nnnnnnnnuc cccuauuuunnn nnnnnnucua uggacgcuu cuugggga 228

<210> SEQ ID NO 368
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Vibrio cholerae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 16-223
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 368

uuccgcogua gaggangcog uuaaogmnaaa ngamuaac ucugunuunnn nnnnggggug 60
auuggcaaiug nnnnnnnnnn uggannnaaa ggnnnguggc gcggcaaguc cccuugcnnnc 120
caucnncng canauggcgg ccggguuaauun nnccnnacau ggguuacgc uggacauagun 180
nnnnunnnuau uguuuauuuu nnnnnnnnnu uggagcogcu cunugug 228

<210> SEQ ID NO 369
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Vibrio cholerae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 7-207
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 369

cuuaaangua ggcgcngcgc uguuuaacu nguucmaaa uucguunnn nnnnagggguug 60
acccgaugnn nnnnnnaauc cuuggnnuuaa ggnngucac gcgcgaagug aucguunuun 120
cucnnnnnc aaccuguucgu ggcgccagcu uunngacauu anucggcgc uggcuauggn 180
nnnnnnnnuuggg unnnnnnnu uuggacgcu cuuauag 228

<210> SEQ ID NO 370
<211> LENGTH: 228
<212> TYPE: RNA
<213> ORGANISM: Vibrio vulnificus
<220>FEATURES:
<221>NAME/KEY: misc_feature
<222>LOCATION: 16-204
<223>OTHER INFORMATION: n = g, a, c or u

<400>SEQUENCE: 370

uuuaccgcaaa gggagagcc ugrnncccgag cggnaauguu uuggggannnn nnngenccgca 60
acccacnnnn nnnnnnnnnc aagcaccuuu gggggagug uggagggua gacaaaaauu 120
ngacccauuu uagacccugc guggacguu gggauguunc ccaannucaac uccuaccgc 180
nnnnnnnnnn uccnnnnnnn nnngenccuga uagaagccuu cugagag 228

<210>SEQ ID NO 371
<211>LENGTH: 228
<212>TYPE: RNA
<213>ORGANISM: Vibrio vulnificus
<220>FEATURE:
<221>NAME/KEY: misc_feature
<222>LOCATION: 16-206
<223>OTHER INFORMATION: n = g, a, c or u

<400>SEQUENCE: 371

uuuacccgcaaa gggccgac uggguuuaagc aguunnnauuuu auaauuunnm nnngenagggug 60
augcccaaugn nnnnnnnnnuu uagunnnnnu ugguggccau uggggagag uaaungcnnn 120
uuuacccauuu uagacccugc guggacguu gggauguunc ccaannucaac uccuaccgc 180
nnnnnnnnn uccnnnnnnn nnngenccuga uagaagccuu cugagag 228

<210>SEQ ID NO 372
<211>LENGTH: 486
<212>TYPE: DNA
<213>ORGANISM: Artificial Sequence
<220>FEATURE:
<223>OTHER INFORMATION: Description of Artificial Sequence: Note=Synthetic construct

<400>SEQUENCE: 372

nnnnnnnncccc ttcccaagc nnngenccagc ggaggacygc gnnnnccnnnn gnnnnccnnnc 60
rgccacnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120
nnnnnnnnnnn nnnnnnncg gnnnnnccncc gnnnnncnnnn gnnnnnnnnnnn 180
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 240
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 300
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 360
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 420
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 480
nnnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnnnnnnnnn 486
cagagancnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn cnggggra 238

<210> SEQ ID NO 376
<211> LENGTH: 221
<212> TYPE: DNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 25
<223> OTHER INFORMATION: k = g or t/u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 7-217
<223> OTHER INFORMATION: n = g, a, c or t/u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 24, 78, 79, 81, 96, 97, 213
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 153
<223> OTHER INFORMATION: v = g, c or a
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 214, 220
<223> OTHER INFORMATION: w = a or t/u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 169, 221
<223> OTHER INFORMATION: y = c or t/u

<400> SEQUENCE: 376
wagagngcnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 60
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn 120
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnyc gtcnnnnnn 180
nnnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnnnnn tgrwmnctw y 221

<210> SEQ ID NO 377
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = Synthetic construct
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1
<223> OTHER INFORMATION: y = c or u
<400> SEQUENCE: 381

ygpgaaggn 9

<210> SEQ ID NO 382
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-3, 9, 13, 17
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 4, 11
<223> OTHER INFORMATION: r = a or g
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 7
<223> OTHER INFORMATION: y = c or u
<400> SEQUENCE: 382

nnmracygcn ranacnggcc 20

<210> SEQ ID NO 383
<211> LENGTH: 6
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 2-9, 15-19
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 383
cugaga 6

<210> SEQ ID NO 384
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 2-9, 15-19
<223> OTHER INFORMATION: n = g, a, c or u
<400> SEQUENCE: 384

amnnnnnnn cugnnmnc 20

<210> SEQ ID NO 385
<211> LENGTH: 19
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 14
<223> OTHER INFORMATION: d = g, a, or u
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 2-7, 9-11
<223> OTHER INFORMATION: n = g, a, c or u

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 10
<223> OTHER INFORMATION: r = a or g

<400> SEQUENCE: 385
unmnnnnn ngdaggra 19

<210> SEQ ID NO 386
<211> LENGTH: 9
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: /Note = synthetic construct

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 9
<223> OTHER INFORMATION: n = g, a, c or u

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 6
<223> OTHER INFORMATION: r = a or g

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 3, 7
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 386
agycrygn 9

<210> SEQ ID NO 387
<211> LENGTH: 50
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: /Note = synthetic construct

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 10, 15
<223> OTHER INFORMATION: k = g or u

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 11, 14, 30-32
<223> OTHER INFORMATION: n = g, a, c or u

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 7, 12, 18-21, 27, 43-44, 48-50
<223> OTHER INFORMATION: r = a or g

<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 4-6, 17, 37
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 387
ngayyrygk nrankcyrxrrocgaacrgun nngaucygyagugrragarr 50

<210> SEQ ID NO 388
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURES:
<223> OTHER INFORMATION: Description of Artificial Sequence: /Note = synthetic construct
<220> FEATURES:
-continued

<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2, 9-10, 13-16, 18
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 17
<223> OTHER INFORMATION: r = a or g
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 7
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 389

ngngcycyann cmnnnrrn

<210> SEQ ID NO 389
<211> LENGTH: 14
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 1, 3-4, 6-7, 14
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 5, 11
<223> OTHER INFORMATION: r = a or g
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 2
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 389

nmnmnmgau ragn

<210> SEQ ID NO 390
<211> LENGTH: 3
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct

<400> SEQUENCE: 390

gag

<210> SEQ ID NO 391
<211> LENGTH: 2
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE: 
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 391

nn

<210> SEQ ID NO 392
<211> LENGTH: 2
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 392

nn 2

<210> SEQ ID NO 393
<211> LENGTH: 44
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-8, 14-20, 21-22, 32-44
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 9-10, 29
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 23, 31
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 393

nnnnnnnn agnnnnnnnn nynygcarg ynnnnnnnn nn

<210> SEQ ID NO 394
<211> LENGTH: 26
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2, 18-28
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 13
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 14
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 394

nnnnnnnn nnyysggnn nn

<210> SEQ ID NO 395
<211> LENGTH: 2
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct

<400> SEQUENCE: 395

aa 2
<210> SEQ ID NO 396
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-11
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 12
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 396

nnnnnnnnn nyuguca 17

<210> SEQ ID NO 397
<211> LENGTH: 11
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:/Note = synthetic construct
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 6
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 7
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 10
<223> OTHER INFORMATION: w = a or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 11
<223> OTHER INFORMATION: y = c or u

<400> SEQUENCE: 397

ugagagnouc y 11

<210> SEQ ID NO 398
<211> LENGTH: 20
<212> TYPE: RNA
<213> ORGANISM: Arabidopsis thaliana
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 2-9, 17-19
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 398

amnnnnnnnn ccgauunnnng 20

<210> SEQ ID NO 399
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Arabidopsis thaliana
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 14
<223> OTHER INFORMATION: d = g, a, or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 2-7, 9-11, 20-22
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURES:
<221> NAME/KEY: misc_feature
<222> LOCATION: 18
<223> OTHER INFORMATION: r = a or g

<400> SEQUENCE: 399

nnnnnnnn nogdaggran nn 22

<210> SEQ ID NO 400
<211> LENGTH: 7
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-7
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 400

nnnnnn 7

<210> SEQ ID NO 401
<211> LENGTH: 3
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 401

gag 3

<210> SEQ ID NO 402
<211> LENGTH: 2
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 402

nn 2

<210> SEQ ID NO 403
<211> LENGTH: 2
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-2
<223> OTHER INFORMATION: n = g, a, c or u

<400> SEQUENCE: 403

nn 2

<210> SEQ ID NO 404
<211> LENGTH: 38
<212> TYPE: RNA
<213> ORGANISM: Bacillus subtilis
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 1-8, 14-20, 30-38
<223> OTHER INFORMATION: n = g, a, c or u
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 9-10, 27
<223> OTHER INFORMATION: r = a or g
<220> FEATURE:
<221> NAME/KEY: misc_feature
18. A method comprising
(a) testing a compound for inhibition of gene expression of a gene encoding an RNA comprising a riboswitch, wherein the inhibition is via the riboswitch,
(b) inhibiting gene expression by bringing into contact a cell and a compound that inhibited gene expression in step (a), wherein the cell comprises a gene encoding an RNA comprising a riboswitch, wherein the compound inhibits expression of the gene by binding to the riboswitch.

19. (canceled)

20. The method of claim 18, wherein the cell is killed or growth of the cell is inhibited.

21. The method of claim 18, wherein the cell is a bacterial cell.

22. The method of claim 18, wherein the cell is in a patient.

23. The method of claim 18, wherein the cell is a bacterial cell, wherein the cell is in a patient.

24. The method of claim 18, wherein the riboswitch is a guanine-responsive riboswitch.

25. The method of claim 18, wherein the riboswitch is an adenine-responsive riboswitch.

26. The method of claim 18, wherein the riboswitch is a lysine-responsive riboswitch.

27. The method of claim 18, wherein the riboswitch is a thiamine pyrophosphate-responsive riboswitch.

28. The method of claim 18, wherein the riboswitch is a flavin mononucleotide-responsive riboswitch.

29. The method of claim 18, wherein the riboswitch is an S-adenosylmethionine-responsive riboswitch.

30. A method comprising
(a) testing a compound for inhibition of expression of a gene encoding an RNA comprising a riboswitch, wherein the inhibition is via the riboswitch, wherein the compound inhibits expression of the gene encoding the RNA comprising the riboswitch then the compound is identified as a trigger molecule of the riboswitch, wherein the compound inhibits expression of the gene by binding to the riboswitch.

31. The method of claim 30, wherein the riboswitch is a guanine-responsive riboswitch.
32. The method of claim 30, wherein the riboswitch is an adenine-responsive riboswitch.

33. The method of claim 30, wherein the riboswitch is a lysine-responsive riboswitch.

34. The method of claim 30, wherein the riboswitch is a thiamine pyrophosphate-responsive riboswitch.

35. The method of claim 30, wherein the riboswitch is a flavin mononucleotide-responsive riboswitch.

36. The method of claim 30, wherein the riboswitch is an S-adenosylmethionine-responsive riboswitch.

* * * * *