METHODS AND COMPOSITIONS FOR REGULATING GENE EXPRESSION IN PLANT CELLS

Inventors: Roger N. Beachy, St. Louis, MO (US); Maria Isabel Ordiz Luis, St. Louis, MO (US); Shunhong Dai, St. Louis, MO (US)

Assignee: Donald Danforth Plant Science Center, St. Louis, MO (US)

Abstract

Novel chimeric plant promoter sequences are provided, together with plant gene expression cassettes comprising such sequences. In certain preferred embodiments, the chimeric plant promoters comprise the BoxII cis element and/or derivatives thereof. In addition, novel transcription factors are provided, together with nucleic acid sequences encoding such transcription factors and plant gene expression cassettes comprising such nucleic acid sequences. In certain preferred embodiments, the novel transcription factors comprise the acidic domain, or fragments thereof, of the RF2a transcription factor. Methods for using the chimeric plant promoters and novel transcription factors in regulating the expression of at least one gene of interest are provided, together with transgenic plants comprising such chimeric plant promoter sequences and novel transcription factors.

14 Claims, 29 Drawing Sheets
Figure 1
Figure 2
Figure 4
Figure 5
Box II  
5' CCAGTGTGCCCCCTGG 3'

Box IIm1  
5' CCAGTGTGCGGCTGG 3'

Box IIm2  
5' GGAGTGTGCCCCCTTC 3'

Figure 6
Figure 7
<table>
<thead>
<tr>
<th></th>
<th>RF2a</th>
<th></th>
<th>RF2b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ka (1/M)</td>
<td>Kd(M)</td>
<td>Ka (1/M)</td>
</tr>
<tr>
<td>Box II</td>
<td>8.67 x 10^6</td>
<td>1.15 x 10^-7</td>
<td>2.19 x 10^6</td>
</tr>
<tr>
<td>Box II m1</td>
<td>1.08 x 10^7</td>
<td>9.26 x 10^-8</td>
<td>4.17 x 10^6</td>
</tr>
<tr>
<td>Box II m2</td>
<td>8.13 x 10^8</td>
<td>1.23 x 10^-7</td>
<td>1.85 x 10^6</td>
</tr>
</tbody>
</table>

**Figure 8**
Figure 9
Figure 10
Box II

-48  TATA  +1  +8  p-1.0hBoxII-48CaMV
-48  TATA  +1  +8  p-48CaMV

Box II

-53  TATA  +1  +8  p-1.5hBoxII-53CaMV
-53  TATA  +1  +8  p-53CaMV

Box II

-69  TATA  +1  +8  p-3.0hBoxII-69CaMV
-69  TATA  +1  +8  p-69CaMV

Box II

-74  TATA  +1  +8  p-3.5hBoxII-74CaMV
-74  TATA  +1  +8  p-74CaMV

Box II

-90  TATA  +1  +8  p-5.0hBoxII-90CaMV
-90  TATA  +1  +8  p-90CaMV

Box II

-95  TATA  +1  +8  p-5.5hBoxII-95CaMV
-95  TATA  +1  +8  p-95CaMV

Figure 11
-1.0hBoxII-48CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGtgcgaagacctttcctc

-1.5hBoxII-53CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGatccttgcgaagacct

-1.5hBoxII-53CaMV-c: 
tgatcaaaagcttatcccttgcgaagacct

-3.0hBoxII-69CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGgaagcacaatccacttaca

-3.5hBoxII-74CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGggtacgcacaatcc

-3.5hBoxII-74CaMV-c: 
tgatcaaaagcttggtacgcacaatcc

-5.0hBoxII-90CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGatctccactgacgtaag

-5.5hBoxII-95CaMV: 
tgatcaaaagcttCCAGTGTCGCCCTGGgtgatctccactgac

-5.5hBoxII-95CaMV-c: 
tgatcaaaagccttggtgatctccactgac

Figure 12
Figure 13
Figure 14
A

-32::GUS

-68::GUS

-100::GUS

E(IIm1)::GUS

E(IIm2)::GUS

E::GUS

B

Figure 15
Figure 16

A

<table>
<thead>
<tr>
<th>RF2a</th>
<th>ΔP</th>
<th>ΔQ</th>
<th>ΔPΔA</th>
<th>ΔPΔQ</th>
<th>3Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b-Zip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

MW

- 70 kDa
- 60 kDa
- 50 kDa
- 40 kDa
- 30 kDa
- 25 kDa
- 20 kDa

C

Free | ΔP | ΔQ | ΔPΔA | ΔPΔQ | 3Δ |

Figure 16
### Figure 17

#### A

**Reporter**
- pBH-48Ca:GUS

**Effector**
- pMON999
- p35S::RF2a
- p35S::RF2a-ΔP
- p35S::RF2a-ΔQ
- p35S::RF2a-ΔPΔA
- p35S::RF2a-3Δ

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Effector</th>
<th>Relative GUS Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBH-48Ca:GUS</td>
<td>pMON999</td>
<td>100%</td>
</tr>
<tr>
<td>pBH-48Ca:GUS</td>
<td>p35S::RF2a</td>
<td>50%</td>
</tr>
<tr>
<td>pBH-48Ca:GUS</td>
<td>p35S::RF2a-ΔP</td>
<td>25%</td>
</tr>
<tr>
<td>pBH-48Ca:GUS</td>
<td>p35S::RF2a-ΔQ</td>
<td>12.5%</td>
</tr>
<tr>
<td>pBH-48Ca:GUS</td>
<td>p35S::RF2a-ΔPΔA</td>
<td>5%</td>
</tr>
<tr>
<td>pBH-48Ca:GUS</td>
<td>p35S::RF2a-3Δ</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

#### B

Graph showing relative GUS activity levels for different combinations of reporters and effectors.
A

Reporter

pC7er2:GUS

Effector

p35S:2C7
p35S:2C7-A
p35S:2C7-P
p35S:2C7-Q
p35S:2C7-PA
p35S:A-2C7
p35S:P-2C7
p35S:Q-2C7

B

Figure 18
<table>
<thead>
<tr>
<th>Plant Lines</th>
<th>Wt</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal Phenotype</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+++</td>
</tr>
</tbody>
</table>

- **Figure 20**
Figure 21
Figure 22
Figure 24
Figure 25
Figure 26
<table>
<thead>
<tr>
<th>Construct</th>
<th>Components</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>5XG-35S:GUS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGE</td>
<td>CsVMV</td>
<td>GE</td>
</tr>
<tr>
<td>ASGE</td>
<td>CsVMV</td>
<td>A5</td>
</tr>
<tr>
<td>A5x2GE</td>
<td>CsVMV</td>
<td>A5</td>
</tr>
<tr>
<td>A5x3GE</td>
<td>CsVMV</td>
<td>A5</td>
</tr>
<tr>
<td>GEA5</td>
<td>CsVMV</td>
<td>GE</td>
</tr>
<tr>
<td>GEA5x2</td>
<td>CsVMV</td>
<td>GE</td>
</tr>
<tr>
<td>GEA5x3</td>
<td>CsVMV</td>
<td>GE</td>
</tr>
<tr>
<td>GEA5x4</td>
<td>CsVMV</td>
<td>GE</td>
</tr>
</tbody>
</table>

**Figure 27**

**A**

**B**

Relative GUS Activity
Figure 29

Bar graphs showing relative GUS activity at different ages (AGE#11, AGE#18) and different treatments (YGE#11, ASASGE#18). The x-axis represents different treatments or conditions, and the y-axis represents relative GUS activity. The graphs illustrate the variation in GUS activity across different treatments and conditions.
METHODS AND COMPOSITIONS FOR
REGULATING GENE EXPRESSION IN
PLANT CELLS

CROSS-REFERENCE TO RELATED
APPLICATIONS

This application is a continuation-in-part and claims the
benefit of U.S. patent application Ser. No. 10/888,613, filed
Jul. 9, 2004, now U.S. Pat. No. 7,517,689 which is incorpo-
rated herein by reference.

This invention was supported in part by grant number
NNA04CC29G from the National Aeronautics and Space
Administration (NASA). The U.S. Government may have
certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates to the field of molecular biol-
y. In particular, the invention relates to methods and com-
positions that can be used to regulate gene expression. Still
more particularly, the invention relates to methods and com-
positions that can be used to initiate and/or enhance the
expression of at least one gene of interest in plant cells.

BACKGROUND OF THE INVENTION

Gene expression may be regulated in several ways, which
include the activation or suppression of transcription, the
differential processing and stabilization of messenger RNA
("mRNA") and the extent of translation of the mRNA. The
control of transcription plays a particularly critical role in
the regulation of gene expression in eukaryotic cells. There are
several structural elements that are involved in the regulation
of transcription.

Promoters represent a class of nucleic acid structures that
are involved in the regulation of transcription. In general,
promoters are located next to the transcription start site and
interact with RNA polymerase, either directly or indirectly.
Promoters often comprise several discrete "cis elements," each
of which may be recognized by one or more trans-acting
regulatory proteins known as transcription factors. Among
the various cis elements well-known in the art is the "TATA
box," which is known to interact with certain regulatory pro-
en, e.g., transcription factors, and is generally located about
20-30 base pairs upstream from the transcription initiation
site.

The binding of such transcription factors to promoters or
other regulatory sequences is often responsible for the initia-
tion, maintenance and/or down-regulation of transcription. A
typical gene-specific eukaryotic transcription factor includes
a DNA-binding domain and one or more additional domains
that influence the activation or repression of transcription, e.g.,
"trans-acting domains." Transcription factors bind in the
general proximity (although occasionally at great distances)
of the point of transcription initiation of a gene. Such tran-
scription factors often act to influence the efficiency of for-
mation or function of a transcription initiation complex at the
promoter. Transcription factors can act in a positive fashion
(transactivation) or in a negative fashion (transrepression).
Furthermore, the effect that transcription factors may have on
gene expression can be constitutive (always "on") or condi-
tional.

Over the years, several classes of DNA-binding domains of
various transcription factors have been characterized and the
nucleic acid sequences to which such domains interact iden-
tified. Non-limiting examples of such domains include motifs
known as the leucine zipper, the bZIP domain, the zinc-finger,
the homeobox, the basic helix-loop-helix and others. The trans-
acting domains of transcription factors are often charac-
terized as having a high content of specific amino acids,
which include domains rich in acidic amino acids, proline or
glutamine (Ginger et al., 1985; Meshi and Iwabuchi, 1995;
Mitchell and Tjian, 1989). Acidic domains have been
reported to possess activation functions that include interac-
tions with TATA-binding proteins ("TBP") (Truant et al.,
1993), TBP-associated factors ("TAFs") (Uesugi et al.,
1997), TFIIA (Pugh, 2000), TFIIIB (Klemm et al., 1995) and
other general transcription complexes (Stargell and Struhl,
1995).

Beachy, in U.S. Pat. No. 5,824,857 entitled "Plant Pro-
moter," described the promoter from the rice tungro bacilli-
form virus ("RTBV"). The '857 patent discloses that the
RTBV promoter causes preferential gene expression in plant
vascular tissue. The patent also discloses that the RTBV pro-
moter can be used to drive expression in most plants, whether
monocotyledonous or dicotyledonous, and is particularly
suited to rice. The patent further discloses the transformation
of plants by inserting the coding sequence of the RTBV
promoter and a heterologous gene of interest to obtain trans-
genic plants that express the gene of interest in vascular
tissue.

Yin and Beachy, in "The regulatory regions of the rice
tungro bacilliform virus promoter and interacting nuclear
factors in rice (Oryza sativa L.), The Plant Journal, 7(6):
969-980 (1995)," described the E fragment (~164 to +45 in
relation to the transcription start site) within the RTBV pro-
moter, which was shown to be sufficient to cause tissue-
specific gene expression. The article also disclosed a critical
cis element, Box II (~53 to ~39), within the E fragment that
was shown to be essential for promoter activity. The same
authors identified other cis elements of the RTBV promoter in
"Promoter elements required for phloem-specific gene
expression from the RTBV promoter in rice, The Plant Jour-
nal: 12(5): 1179-1188 (1997)," including the ASL Box (~98
to ~79) and a GAIA motif (~143 to ~135). Together, these cis
elements were shown to confer phloem-specific reporter gene
expression.

Yin et al., in "RF2a, a bZIP transcriptional activator of
phloem-specific rice tungro bacilliform virus promoter, func-
tion in vascular development, The EMBO Journal, 16(17):
5247-5259 (1997)," identified a 1.8 Kb transcription factor
consisting of 368 amino acids—designated as RF2a. The
RF2a transcription factor is currently known to represent a
bZIP transcription activator found in rice plants that contains
acidic, proline-rich and glutamine-rich putative functional
domains. RF2a has been shown to bind to the Box II element
of the RTBV promoter and stimulate Box II-dependent tran-
scription in vitro. Another bZIP protein, RF2b, has been
isolated through interaction with RF2a, which also has been
shown to interact with the Box II element.

The inventors have discovered that the Box II cis-element
of the RTBV promoter is portable and that it can be used to
modulate gene expression in unrelated promoters in connec-
tion with RF2a and/or RF2b. That is, until now, it was not
known that the Box II element, and similar sequences, could
be used in chimeric promoters to regulate gene expression in
connection with RF2a and/or RF2b. What's more, the inven-
tors have discovered that the acidic domain of RF2a is par-
ticularly critical to the activation function of this transcription
factor and, moreover, that it can be transferred to unrelated
DNA-binding proteins to modulate gene expression. Accord-
ingly, the inventors have discovered a new system for modu-
larizing gene expression, as described further below.
SUMMARY OF THE INVENTION

Until now, it was not known that the Box II element of the RTBV promoter is portable and that it can be used in connection with unrelated promoters to activate and/or enhance gene expression in the presence of RF2a and/or RF2b. Similarly, the inventors have discovered that operational derivatives of the Box II element can be used to regulate gene expression in like fashion. Still further, the inventors have discovered that the acidic domain of RF2a is also portable and that it can be transferred to unrelated DNA-binding proteins to activate and/or enhance gene expression.

Accordingly, the present invention exploits the portability of the Box II sequence, its operational derivatives and its interaction with RF2a and/or RF2b to provide novel compositions and methods that can be used to control the expression of one or more genes of interest. Further, the present invention provides novel transcription factors, which contain at least one domain that comprises an amino acid sequence that is at least substantially similar to the acidic domain of RF2a, which can be used in connection with unrelated DNA-binding domains to regulate gene expression.

In one preferred embodiment, the invention provides novel chimeric promoters, which comprise (a) nucleic acid sequences derived from any promoter (other than the RTBV promoter), or promoter fragment, that are capable of driving gene expression in plant cells and (b) at least one nucleic acid sequence consisting of (i) SEQ ID NO:1; (ii) SEQ ID NO:2; (iii) SEQ ID NO:3; and (iv) sequences that are substantially similar to either SEQ ID NO:1, 2, or 3. The nucleic acid sequence consisting of SEQ ID NO:1, 2, 3 or sequences substantially similar to any of the foregoing sequences, is, preferably, located in a position that is approximately 7 nucleotides from the TATA box of the promoter, or promoter fragment, plus one or more full turns of DNA helix.

In another preferred embodiment, the invention provides plant gene expression cassettes that comprise a first chimeric promoter, which comprises a nucleic acid sequence consisting of (i) SEQ ID NO:1; (ii) SEQ ID NO:2; (iii) SEQ ID NO:3; and (iv) sequences that are substantially similar to either SEQ ID NO:1, 2, or 3. The first chimeric promoter is operatively linked to any gene of interest. The nucleic acid sequence consisting of SEQ ID NO:1, 2, 3 or sequences substantially similar to any of the foregoing sequences, is, preferably, located in a position that is approximately 7 nucleotides from the TATA box of the promoter, or promoter fragment, plus one or more full turns of DNA helix.

In a related embodiment, the invention provides plant gene expression cassettes comprising (i) a first promoter, which comprises a nucleic acid sequence that is capable of interacting with at least one DNA-binding domain of at least one polypeptide, operatively linked to a gene of interest and (ii) a second promoter operatively linked to a nucleic acid sequence that encodes a polypeptide, which comprises an amino acid sequence that is at least substantially similar to SEQ ID NO:6 and a DNA-binding domain that is capable of interacting with the corresponding nucleic acid sequence of the first promoter. In a related embodiment, the present invention provides that the expression of the second promoter may be chemically-induced.

In other embodiments, novel transcription factors and uses thereof are provided. In particular, the inventors have discovered that the acidic domain of RF2a (SEQ ID NO:6), and sequences substantially similar to SEQ ID NO:6, can be fused to unrelated transcription factors, or transcription factor-like complexes, to regulate gene expression. More specifically, the invention provides that novel transcription factors, which comprise the acidic domain of RF2a and at least one DNA-binding domain, can be used to modulate the expression of one or more genes of interest, which are driven by promoters that comprise nucleic acid sequences recognized by at least one DNA-binding domain of such novel transcription factors.

In other embodiments, the invention provides methods of regulating the expression level of at least one gene of interest comprising transforming a plant cell with at least one plant gene expression cassette of the present invention. Still further, the invention provides methods of regulating the expression level of at least one gene of interest comprising (a) transforming plant cells with at least one plant gene expression cassette of the present invention, wherein the cassette comprises a chemically-inducible promoter operatively linked to a nucleic acid sequence that encodes a polypeptide comprising (i) an amino acid sequence that is at least substantially similar to SEQ ID NO:6 and (ii) an amino acid sequence that is capable of interacting with another promoter operatively linked to a gene of interest, wherein the interaction initiates and/or enhances the expression of the gene of interest and (b) contacting plant cells, embryos, tissues, roots, etc., directly or indirectly, derived from the transformed plant cells with an activating amount of the expression-inducing chemical.

In other embodiments, the invention provides that two or more chimeric promoters containing the Box II element and/or its operational derivatives, which are operatively linked to one or more genes of interest, can be used in connection with RF2a- and/or RF2b-encoding sequences to achieve a “cascade-type” system. In this embodiment, the expression of RF2a and/or RF2b activates and/or enhances the expression of the nucleic acid sequences to which the one or more chimeric promoters are operatively linked. Still further, the nucleic acid sequence encoding the RF2a and/or RF2b protein may be operatively linked to a chemically-inducible promoter. In such embodiment, upon contacting plant cells, embryos, or tissues, which have been transformed with such nucleic acid sequences, with the expression-inducing chemical, the RF2a and/or RF2b transcription factor is produced and subsequently interacts with the Box II element and/or its operational derivatives to regulate gene expression. This interaction, of course, results in the synchronized activation or enhancement of expression of all Box II-dependent genes (or all genes operatively linked to promoters containing operational derivatives of Box II).

In other embodiments, the invention provides that at least one chimeric promoter containing the Box II element and/or its operational derivatives, which is operatively linked to one or more genes of interest, can be used in combination with RF2a- and/or RF2b-encoding sequences to “turn-off” or modulate the expression of one or more unrelated endog-
The functions of DNA that can be affected in this embodiment, for example, include replication and transcription. The functions of RNA that can be affected include translational RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be imparted or facilitated by the RNA. The effect of such interference with target nucleic acid function may provide the ability to modulate the expression of particular gene products.

In yet a further embodiment, the invention provides plant cells, plant embryos, plant tissues, whole plants and seeds that have been transformed with at least one plant gene expression cassette of the present invention.

The above-mentioned and additional features of the present invention are further illustrated in the Detailed Description contained herein. All references disclosed herein, including U.S. patents, are hereby incorporated by reference in their entirety as if each was incorporated individually.

DESCRIPTION OF THE FIGURES

FIG. 1: Comparison of RTBV promoter activity with different constitutive promoters in BY-2 protoplasts. "RTBV" represents the promoter of RTBV; "E" represents the E fragment of the RTBV promoter; "35S" refers to the enhanced 35S promoter of CaMV; "35S (-46)" refers to the 5' deletion mutation of the CaMV 35S promoter, which ends at position −46 as described herein; "CaMV" represents the promoter of CaMV. The results presented are the mean value of three independent experiments with standard variations. Each experiment included three repeats and all data were normalized against GUS as internal control.

FIG. 2: Relative GUS activity of each plasmid driven by truncated mutants of the RTBV promoter. The results presented are the mean value of three independent experiments. Each experiment includes three repeats and all data were normalized against GUS as internal control.

FIG. 3: Diagram of constructs that were prepared with the Box II element inserted into different locations in the context of the E fragment of the RTBV promoter. The modified promoters were then inserted into a vector, wherein the modified promoters were operatively linked to the uidA coding sequence followed by a Nos terminator. The names of the derived plasmids are shown to the right side of each diagram.

FIG. 4: Activities of the modified promoters shown in FIG. 3. The results presented are the mean value of three independent experiments with standard variation. Each experiment included three repeats and all data were normalized against GUS as internal control.

FIG. 5: E fragment of the RTBV promoter was activated in tobacco BY-2 protoplasts by co-transfection of E:GUS with CaMV::RF2b; CaMV 35S::RF2a and CaMV 35S::RF2a/CaMV::RF2b. The modified promoters were then inserted into a vector, wherein the modified promoters were operatively linked to the uidA coding sequence followed by a Nos terminator.

FIG. 6: Sequences of wild type Box II (labeled Box II) and two non-limiting examples of its operational derivatives (labeled as Box Iml1 and Box Iml2). The mutated nucleotides in each operational derivative are identified in bold underline.

FIG. 7: Relative GUS activity of constructs p(E(Box IIml1):GUS, p(E(Box Iml1)):GUS, p(E(Box Iml2)):GUS and p(E::GUS in BY-2 protoplasts as described herein. p(E(Box Iml1)) comprised the E fragment, wherein the wild type Box II was replaced with the Box Iml1 element; p(E(Box Iml2)) comprised the E fragment, wherein the wild type Box II was replaced with the Box Iml2 element; p(E::GUS comprised the E fragment and the wild type Box II element; and the bar labeled "BY-2" represents non-transgenic protoplasts.

FIG. 8: Comparative analysis of DNA binding affinities of RF2a and RF2b.

FIG. 9: The p(E(Box Iml1)):GUS construct was co-transfected into tobacco BY-2 protoplasts with CsVMV::RF2b, CaMV 35S::RF2a and 35S::RF2a/CaMV::RF2b. All samples were normalized with GFP internal control.

FIG. 10: The p(E(Box Iml2)):GUS construct was co-transfected into tobacco BY-2 protoplasts with CsVMV::RF2b, CaMV 35S::RF2a and CaMV 35S::RF2a/CaMV::RF2b. All samples were normalized with GFP internal control.

FIG. 11: Diagram of Box II and CaMV 35S chimeric promoters (and control). The chimeric promoters were ligated with a plasmid cassette that contained the uidA coding sequence followed by a Nos terminator. The names of the derived plasmids are shown to the right side of the diagram. The positions of Box II in the various constructs are in relation to the site of transcription initiation (+1).

FIG. 12: Primers used in generating the chimeric promoters containing the Box II element and portions of the CaMV 35S promoter, which are illustrated in FIG. 11. To facilitate the cloning process, a HindIII restriction site was added to the 5' end of the primers (underlined). Box II elements are shown in bold and upper case letters. The sequences from the CaMV 35S promoter are shown in lower case italic letters.

FIG. 13: Relative GUS activities in non-transgenic BY-2 protoplasts of constructs in which uidA was driven by promoters that comprised Box II and portions of the CaMV 35S promoter, which are illustrated in FIG. 11. The GUS activity of each sample was normalized against the GFP internal control.

FIG. 14: A: Relative GUS activities of constructs in which the uidA gene was driven by Box II and CaMV 35S fusion promoters (which are illustrated in FIG. 11), wherein the constructs were inserted into BY-2 protoplasts that produce RF2a. B: Relative GUS activities of the constructs referenced in (A) above, wherein the constructs were inserted into BY-2 protoplasts that produce RF2a and RF2b. The GUS activity of each sample was normalized against the GFP internal control.

FIG. 15: A: Diagram of the T-DNA regions of binary plasmids used for Agrobacterium-mediated transformation of Arabidopsis thaliana. B: GUS activity of T1 transgenic Arabidopsis plants. The results present the mean value of at least 15 independent transgenic plants. The relative GUS activity was calculated by comparison with E::GUS (=1).

FIG. 16: Electrophoretic mobility shift assay of protein-DNA complexes formed between mutants of RF2a and the Box II element. A: Schematic diagram of mutants of RF2a. "A" represents the acidic domain; "P" represents the proline-rich domain; and "Q" represents the glutamine-rich domain. B: SDS-PAGE analysis of the purified RF2a mutant proteins. C: Gel mobility shift assay using purified mutant proteins of RF2a as labeled. A control lane without protein (labeled "Free") was included in the assay. Box Iml1 DNA was labeled with [32P] and radioactivity was detected by autoradiography.

FIG. 17: Effects of RF2a and mutants of RF2a on gene expression in BY-2 protoplasts. A: Reporter and effector gene constructs. The GUS reporter gene (uidA sequence) was driven by a promoter that comprised the Box II element operatively linked to nucleotides −48 to +8 of the CaMV 35S sequence followed by a Nos terminator.
promoter, and was followed by a nopalin synthase 3' terminator sequence ("pBI-48Ca:GUS"). Effectors included sequences encoding RF2a and RF2a mutants operatively linked to the CaMV 35S promoter and nopalin synthase terminator. B: Relative GUS activities in BY-2 protoplasts that were co-transfected with Reporter and Effector gene constructs described herein. The results represent the averages of three repeats, three plants for each repeat (with standard deviation). Open bars represent control leaf tissue that was not subjected to Intrepid® 2F treatment. Solid bars represent leaf tissue treated with 1:8,000 dilution of Intrepid® 2F. B: Western Blot showing expression of RF2a in a limited number of samples analyzed in FIG.22A. The lower panel shows the SDS-PAGE gel described below, whereas the upper panel shows the detected RF2a protein on the nitrocellulose membrane. The mark "*" refers to control leaf tissue that was not treated with Intrepid® 2F, whereas "**" refers to leaf tissue that was treated with 1:8,000 dilution of Intrepid® 2F. The RF2a-arrow indicates the band position of the RF2a protein.

FIG. 23: A bar graph summarizing the expression-enhancing activity of various fragments of the acidic domain of RF2a, as further described in Example 14.

FIG. 24: A Diagram of the different constructs used to transform BY-2 protoplasts, as further described in Example 15. B: A bar graph showing the relative GUS activity produced by the various constructs transformed into such BY-2 protoplasts, as further described in Example 15 below.

FIG. 25: A diagram of the full-length RF2a protein, showing proline-rich, acidic, bZIP, and glutamine-rich domains.

FIG. 26: A bar graph summarizing the expression-enhancing activity of the acidic domain of RF2a in Arabidopsis thaliana plants, as further described in Example 16.

FIG. 27: A Diagram of the different constructs used to transform BY-2 protoplasts, as further described in Example 17. B: A bar graph showing the relative GUS activity produced by the various constructs transformed into such BY-2 protoplasts, as further described in Example 17 below.

FIG. 28: Analysis of T1 generation Arabidopsis thaliana plants transformed with VGE, AGE, or A5ASGE (and GUS reporter) constructs, as further described in Example 18 below. Top: A bar graph showing the relative GUS activity produced by the various VGE constructs. Middle: A bar graph showing the relative GUS activity produced by the various AGE constructs. Bottom: A bar graph showing the relative GUS activity produced by the various A5ASGE constructs.

FIG. 29: Analysis of T1 generation Arabidopsis thaliana plants transformed with VGE, AGE, or A5ASGE (and GUS reporter) constructs, as further described in Example 18 below. Top: A bar graph showing the relative GUS activity produced by the various VGE constructs. Middle: A bar graph showing the relative GUS activity produced by the various AGE constructs. Bottom: A bar graph showing the relative GUS activity produced by the various A5ASGE constructs.

DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NO. 1: The nucleic acid sequence of the Box II element.

SEQ ID NO. 2: The nucleic acid sequence of the Box IIm1 element—an operational derivative of the Box II element.

SEQ ID NO. 3: The nucleic acid sequence of the Box IIm2 element—an operational derivative of the Box II element.

SEQ ID NO. 4: The amino acid sequence of the RF2a transcription factor.

SEQ ID NO. 5: The amino acid sequence of the RF2b transcription factor.

SEQ ID NO. 6: The amino acid sequence of the acidic domain of the RF2a transcription factor.

SEQ ID NO. 7: Non-limiting example of nucleic acid sequence that encodes the RF2a transcription factor.

SEQ ID NO. 8: Non-limiting example of nucleic acid sequence that encodes the RF2b transcription factor.
The following will describe in detail several preferred embodiments of the invention. These embodiments are provided by way of explanation only, and thus, should not unduly restrict the scope of the invention. In fact, those of ordinary skill in the art will appreciate upon reading the present specification and viewing the present drawings that many variations and modifications of the invention may be employed, used and made without departing from the scope and spirit of the invention.

The present invention can be viewed as having at least two components. The first of the at least two components relates to the Box II element; operational derivatives of the Box II element; chimeric promoters containing one or more of these elements; plant gene expression cassettes containing at least one chimeric promoter of the present invention; plant cells, embryos and tissues transformed with at least one expression cassette of the present invention and methods of using the foregoing compositions in connection with at least one gene of interest. The second of the at least two components relates to (i) novel transcription factors that comprise at least one DNA-binding domain and an amino acid sequence that is at least substantially similar to the acidic domain of RF2a and (ii) methods of using the novel transcription factors to regulate the expression of at least one gene of interest.

Novel Chimeric Promoters

In one preferred embodiment, the invention provides novel chimeric promoters. The promoters of the present invention comprise a first nucleic acid sequence derived from any promoter (other than the RTBV promoter (SEQ ID NO:62)), or promoter fragment, that is capable of driving gene expression in plant cells. The promoters further comprise at least one nucleic acid sequence selected from the group consisting of: (i) the Box II element (SEQ ID NO:1); (ii) the Box IIm1 element (SEQ ID NO:2); (iii) the Box IIml element (SEQ ID NO:3); (iv) sequences that are substantially similar to either the Box II, Box IIm1 or Box IIml element and other nucleic acid sequences that are capable of interacting with RF2a and/or RF2b to regulate gene expression.

It will be understood by those skilled in the art that two nucleic acid sequences are “substantially similar” when approximately 70% or more (preferably at least about 80%, approximately 70% or more, but preferably 80% or more, and most preferably 85% or more) of the nucleotides in each sequence are identical.

SEQ ID NO. 9: Non-limiting example of nucleic acid sequence that encodes the acidic domain of RF2a transcription factor.

SEQ ID NO. 10: Primer sequence of Box II-del-3'.

SEQ ID NO. 11: Primer sequence of Box II-del-5'.
and most preferably at least about 90% or 95%) of the nucleotides match over the defined length of the nucleic acid sequence. Sequences that are substantially homologous can be identified by comparing the sequences using readily accessible computer software, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions for a particular system is within the skill of the art. It will be further understood by those skilled in the art that the phrase “at least substantially similar” refers to sequences that are “substantially similar” as described above or, alternatively, identical to one another.

As used herein, “substantially similar” is further meant to include a nucleic acid sequence which, by virtue of the degeneracy of the genetic code, is not identical with that shown in any of the sequences shown in the Sequence Listing, but which still encodes the same amino acid sequence; or a modified nucleic acid sequence that encodes a different amino acid sequence that retains substantially the same activities of the original proteins, either because one amino acid is replaced with a similar amino acid, or because the change (whether it be substitution, deletion or insertion) does not affect the active site of the protein. Thus, it is contemplated by the inventors that various changes may be made in the nucleic acid sequences disclosed, and, of course, the encoded polypeptides, without appreciable loss of their biological activity or utility in the present invention.

The chimeric promoters of the present invention are preferably constructed by combining the first nucleic acid sequence derived from any plant functional promoter, or promoter fragment, with at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, 2, 3 and sequences substantially similar to SEQ ID NO:1, 2 or 3. More preferably, however, the chimeric promoters of the present invention are constructed by combining the first nucleic acid sequence derived from any plant functional promoter, or promoter fragment, with at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:1, 2 and 3. The term “chimeric promoter,” as used herein, refers to any plant functional promoter sequence, or promoter fragment, that comprises SEQ ID NO:1, 2, 3, or sequences substantially similar thereto, wherein such plant functional promoter sequence is not derived from the RTBV promoter.

The nucleic acid sequences consisting of SEQ ID NO:1, 2, 3, and sequences substantially similar thereto, may be positioned within the first nucleic acid sequence described above, which comprises any plant functional promoter (or promoter fragment). Alternatively, the nucleic acid sequences consisting of SEQ ID NO:1, 2, 3, and sequences that are substantially similar to SEQ ID NO:1, 2 or 3, may be fused to the 5' or 3' end of such plant functional promoter (or promoter fragment).

Although SEQ ID NO:1, 2, 3, and sequences substantially similar thereto, may be positioned within, or fused to the end of, a chimeric promoter, such sequences are, preferably, positioned in specific locations of the DNA helix. In particular, such nucleic acid sequences are, preferably, operably linked to the 5' end of the selected promoter, or promoter fragment, with a space of approximately 7 nucleotides from the TATA box plus one or more full “turns of DNA helix.” It is well known in the art that one turn of DNA helix equals, approximately, 10.4 base pairs. Furthermore, the TATA box, which, generally, is the module within a promoter that functions to position the start site for RNA synthesis, and its location in a promoter can be easily identified by those skilled in the art. In many cases, the TATA box consensus sequence (TATAAT) is 20 to 30 base pairs upstream (i.e., 5') of the transcription start site (by convention ~30 to ~20 base pairs relative to the transcription start site).

The Box II element and its operational derivatives may be used in connection with a variety of promoters, or promoter fragments, to create novel chimeric promoters. Preferably, however, the Box II element and/or its operational derivatives are operably linked to a promoter, or promoter fragment, that comprises a transcription initiation domain. The term “transcription initiation domain” refers to a sequence having at least an RNA polymerase binding site and an mRNA initiation site.

Such basic guidelines for promoter, or promoter fragment, selection emphasize the notion that the Box II element and its operational derivatives are, indeed, “portable” and can be transferred to a plurality of promoter sequences. Of course, the promoter sequence with which the Box II element and/or its operational derivatives will be used, preferably, relates to the target cell-type in which gene expression is desired.

For example, some promoters are known to be active in particular cell-types, in certain tissues, under certain abiotic conditions and/or in the presence of certain inducible agents. Thus, the cell-type and/or conditions in which gene expression is desired will impact the identity of the promoter sequence with which Box II and/or its operational derivatives will be used. Although the endogenous promoter of the gene of interest may be utilized herein for transcriptional regulation, preferably, the promoter is a foreign regulatory sequence. For plant expression vectors, suitable viral promoters include, for example, the RTBV promoter, the 35S RNA and 19S RNA promoters of the cauliflower mosaic virus (“CaMV”); the full-length transcript promoter from frogwort mosaic virus (“FMV”); and the cassava vein mosaic virus (“CSVVM”) promoter.

The Box II element, of course, can be isolated from natural sources, e.g., the Box II element can be isolated from the E fragment of the RTBV promoter (SEQ ID NO:62) using techniques well-known in the art. More preferably, however, the Box II element and its operational derivatives are synthesized using standard DNA synthesis techniques (see, for example, Current Protocols in Molecular Biology, Unit 2.11, eds. Ausubel et al., John Wiley & Sons, 1995).

Expression Cassettes Comprising Novel Chimeric Promoters

The present invention further provides plant gene expression cassettes that comprise a first chimeric promoter of the present invention operatively linked to one or more genes of interest. The expression cassettes may further comprise a second promoter operatively linked to a nucleic acid sequence that encodes a polypeptide, which comprises an amino acid sequence selected from the group consisting of: (a) SEQ ID NO:4; (b) SEQ ID NO:5; (c) SEQ ID NO:6 and (d) sequences that are substantially similar to either SEQ ID NO:4, 5 or 6. Non-limiting examples of nucleic acid sequences that encode the amino acid sequences set forth in SEQ ID NO:4, SEQ ID NO:5 and SEQ ID NO:6 are shown in SEQ ID NO:7, SEQ ID NO:8 and SEQ ID NO:9, respectively.

The one or more genes of interest operatively linked to a first chimeric promoter of the present invention, when including an open reading frame (“ORF”), may encode a protein. Of course, the gene of interest can be an endogenous or exogenous sequence. In addition, the ORF may include certain 5' and 3' untranslated sequences. Still further, appropriate transcription termination and polyadenylation sequences may, preferably, be included.
example, sequences that naturally exist in plants, animals, bacteria, viruses or fungi; synthetic DNA sequences that encode a specific RNA or protein product; cDNA sequences derived from mRNA; DNA sequences modified by mutagenesis, for example, through site specific mutagenesis; chimeras of any of the above (to produce fusion proteins); and DNA sequences encoding complementary RNA molecules (for “antisense” applications); and combinations and/or fragments of the above.

Examples of proteins that can be encoded by the gene of interest include, but are not limited to, nutritionally important proteins; growth promoting factors; proteins for early flowering in plants; proteins that impart protection to the plant under certain environmental conditions, e.g., proteins conferring resistance to metals or other toxic substances, such as herbicides or pesticides; stress-related proteins that confer tolerance to temperature or hydration extremes; proteins conferring resistance to fungi, bacteria, viruses, insects and nematodes; and proteins of specific commercial value, e.g., enzymes involved in metabolic pathways, proteins having therapeutic activity in humans, and others.

The term “operably linked,” as used herein, refers to the functional linkage between, for example, the Box II element and/or its operational derivatives and a plant functional promoter or promoter fragment; between a promoter sequence, including the novel chimeric promoters of the present invention, and any gene of interest; and between a promoter sequence and any sequence encoding the RF2a protein, RF2b protein, a protein comprising the acidic domain of RF2a and/or any protein that is substantially similar to the foregoing proteins.

The second promoter operatively linked to a nucleic acid sequence that encodes a polypeptide, which comprises an amino acid sequence selected from the group consisting of: (a) SEQ ID NO:4; (b) SEQ ID NO:5; (c) SEQ ID NO:6; and (d) sequences that are substantially similar to either SEQ ID NO:4, 5, or 6, may constitute any promoter that is capable of driving gene expression in the target plant cells. As described above, certain promoters are known to be active in particular cell-types, in certain tissues, under certain abiotic conditions and/or in the presence of certain inducible agents. Thus, the cell-type and/or conditions in which gene expression is desired will impact the identity of the promoter that is selected to drive expression of sequences encoding RF2a, RF2b, polypeptides comprising the acidic domain of RF2a or polypeptides that are substantially similar to any of the foregoing. Of course, such promoters may include both constitutive, e.g., the CaMV promoter, and inducible promoters.

By placing the nucleic acid sequence encoding RF2a, RF2b, polypeptides comprising the acidic domain of RF2a or polypeptides that are substantially similar to any of the foregoing, under the control of a chemically-inducible promoter, the gene expression system can be activated at will. The use of inducible promoters in this capacity provides control over the effect that the encoded proteins may have on the expression of promoter sequences (and genes operatively linked to such sequences) comprising the Box II element and/or its operational derivatives. In short, expression of the nucleic acid sequence to which the second promoter is operatively linked is stimulated and/or enhanced by applying an effective amount of the chemical inducer to the plant cells, embryos or tissues that have been transformed with the appropriate expression cassette of the present invention. Application of the inducer, therefore, allows the expressed transcription factor to interact with, and stimulate (or enhance) the expression of, the chimeric promoter and the one or more genes of interest to which it is operatively linked.

To be most useful, an inducible promoter should 1) provide low expression in the absence of the inducer, 2) provide high expression in the presence of the inducer; 3) employ an induction scheme that does not interfere with the normal physiology of the plant; and 4) has no effect on the expression of other genes. Examples of inducible expression schemes useful in plants include those induced by chemical means, such as the ecdysone agonist-inducible gene expression systems (Christopherson et al., 1992; Martinez et al., 1999). The ecdysone agonist-inducible gene expression systems, preferably, employ commercially-available non-steroidal ecdysone agonists, such as tebufenozide and methoxyfenozide.

Additional examples of inducible promoters useful in plants include, but are not limited to, promoters that respond to tetracycline (Gatz et al., 1992; Weinmann et al., 1994); the yeast metallothionein promoter, which is activated by copper ions; the In2-1 and In2-2 regulator sequences, which are activated by substituted benzenesulfonamides, e.g., herbicide safeners; and GRE regulatory sequences, which are induced by glucocorticoids. In addition, plant promoters such as the light-inducible promoter from the small subunit of ribulose bis-phosphate carboxylase (ssRUBISCO); mannopine synthase promoter; nopaline synthase (“NOS”) and octopine synthase (“OCS”) promoters (carried on tumor-inducing plasmids of Agrobacterium tumefaciens) or heat shock promoters, e.g., soybean hsp17.5-E or hsp17.3-B, may be used. Other promoters, both constitutive and inducible, will be known to those of skill in the art.

It will be appreciated by those skilled in the art that the first chimeric promoter, which is operatively linked to at least one gene of interest, and the second promoter, which is operatively linked to a sequence encoding RF2a, RF2b, polypeptides comprising the acidic domain of RF2a or polypeptides that are substantially similar to any of the foregoing, may exist in a single gene expression cassette (or vector), or, alternatively, in separate cassettes (or vectors). Still further, those skilled in the art will appreciate that an expression cassette may comprise a single chimeric promoter, or, alternatively, a plurality of chimeric promoters, each operatively linked to at least one gene of interest. Moreover, the plurality of chimeric promoters may be substantially similar in sequence, or, alternatively, may comprise significantly different sequences.

Novel Transcription Factors

Over the years, several classes of DNA-binding proteins have been characterized and the nucleic acid sequences with which such proteins interact identified. For example, zinc finger proteins represent a class of motifs that are known to be involved in the sequence-specific recognition of DNA. As the name implies, this DNA-binding domain is folded around a zinc ion. To date, more than two hundred proteins, many of them transcription factors, have been shown to possess zinc finger domains. In general, zinc fingers connect transcription factors to their target locations by binding to specific DNA sequences.

The RF2a and RF2b transcription factors described herein represent yet another class of DNA-binding proteins. Specifically, RF2a and RF2b are representative examples of what are commonly referred to as bZIP transcription factors. The bZIP transcription factors are generally characterized by a bipartite DNA-binding domain consisting of a basic region involved in sequence-specific binding, and a leucine zipper region required for dimerization. The bZIP domain of RF2a shares high similarity with bZIP proteins that exist naturally in plants such as Arabidopsis, tobacco, tomato, and other plants (Fukazawa et al., 2000; Jakoby et al., 2002; Ringli and Keller, 1998). In particular, this group of proteins is known to have a
lysine residue at the -10 position relative to the first leucine residue of the leucine zipper region (Yin et al., 1997). The amino acid sequence signature of the DNA-binding regions of this class of proteins is NXXXSAXXXSK (Fuji et al., 2000) (SEQ ID NO: 109).

The present invention provides novel transcription factors, and uses thereof, which provide the ability to regulate the expression of at least one gene of interest. In particular, the inventors have discovered that the acidic domain of RF2a (SEQ ID NO:6), and amino acid sequences substantially similar to SEQ ID NO:6, can be fused to unrelated transcription factors, or transcription factor-like complexes, to regulate gene expression. In certain embodiments, the invention provides that novel transcription factors, which comprise the acidic domain of RF2a and at least one DNA-binding domain, can be used to modulate the expression of one or more genes driven by promoters that encompass nucleic acid sequences recognized by such DNA-binding domain of the novel transcription factors.

In certain embodiments, the acidic domain of RF2a (SEQ ID NO:6), and/or amino acid sequences substantially similar to SEQ ID NO:6, can be fused to any class of DNA-binding domains (and/or unrelated transcription factors comprising such domains). The acidic domain of RF2a, and/or substantially similar sequences, can be fused, for example, to any polypeptide comprising a leucine zipper, bZIP domain, zinc-finger, homeobox, basic helix-loop-helix domain, or other DNA-binding domains currently known in the art (or discovered hereafter). It should be appreciated, however, that the DNA-binding domain selected (or the polypeptide comprising such domain) must be capable of interacting with the promoter operatively linked to the gene of interest for which control of expression is desired.

It will be appreciated by those skilled in the art that two amino acid sequences are “substantially similar” when approximately 70% or more (preferably at least about 80%, and more preferably at least about 90% or 95%) of the amino acids match over the defined length of the sequences. The term “substantially similar” is further meant to refer to amino acid sequences that have been modified from an original sequence, wherein the modified amino acid sequence retains substantially the same level of activity as the original amino acid sequence. This retention in activity, of course, may occur when one or more amino acids are replaced with similar amino acids, or because the change (whether it be substitution, deletion or insertion) does not affect the active site of the protein. Thus, it is contemplated by the inventors that various changes may be made to the acidic domain of RF2a, without appreciable loss of its biological activity and utility in the present invention. It will be further understood by those skilled in the art that the phrase “at least substantially similar” as it is used herein with respect to amino acid sequences refers to sequences that are “substantially similar” as described above or, alternatively, identical to one another.

In certain embodiments, for example, novel transcription factors are provided, which comprise a fragment of the acidic domain of RF2a (SEQ ID NO:6). More particularly, the present invention contemplates that internal regions and fragments of the acidic domain of RF2a (SEQ ID NO:6) may be used to form novel transcription factors, which are capable of regulating gene expression as described herein.

As used herein, “fragments of the acidic domain” and “fragment of the acidic domain” refer to amino acid sequences comprising less than all of the amino acid residues of SEQ ID NO:6, wherein such amino acid sequence is substantially similar to the corresponding region of the full-length SEQ ID NO:6. More particularly, for example, a “fragment of the acidic domain” includes an amino acid sequence encompassing at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of SEQ ID NO:6, wherein such amino acid sequence is substantially similar to the corresponding region of the full-length SEQ ID NO:6.

The acidic domain of RF2a (SEQ ID NO:6) and/or amino acid sequences substantially similar to SEQ ID NO:6 and fragments of the acidic domain may be fused to any class of DNA-binding domains (and/or unrelated transcription factors comprising such domains) to create novel transcription factors as described herein, the invention provides that such sequences may act in the absence of DNA-binding domains to regulate gene expression. In such embodiment, the acidic domain of RF2a may be used alone or in connection with other regulatory proteins or domains to modulate gene expression. For example, the invention provides novel transcription factors that comprise the acidic domain of RF2a and any other amino acid sequence not derived from RF2a or RF2b, which are capable of regulating the expression of at least one gene of interest. The numerous mechanisms by which such transcription factors may regulate expression are well-known to those skilled in the art, which include, for example, affecting the formation of the transcription initiation complex and recruiting other regulatory proteins to the transcription initiation region.

Expression Cassettes Encoding Novel Transcription Factors

The present invention further provides plant gene expression cassettes comprising (i) a first promoter, which comprises a nucleic acid sequence that is capable of interacting with at least one DNA-binding domain of at least one polypeptide, operatively linked to a gene of interest and (ii) a second promoter operatively linked to a nucleic acid sequence that encodes a polypeptide, which comprises an amino acid sequence that is at least substantially similar to SEQ ID NO:6 and a DNA-binding domain that is capable of interacting with the corresponding nucleic acid sequence of the first promoter.

Of course, in light of the foregoing, nucleic acid sequences encoding the acidic domain of RF2a, and/or substantially similar amino acid sequences, may be operatively linked to, for example, nucleic acid sequences that encode polypeptides comprising any motif known to interact with the first promoter (or, more specifically, elements contained within or near the first promoter), which is operatively linked to the one or more genes of interest. This interaction will, preferably, initiate and/or enhance the expression of the one or more genes of interest. An example of a nucleic acid sequence that encodes the acidic domain of RF2a (SEQ ID NO:6) includes, but is not limited to, the sequence shown in SEQ ID NO:9.

Examples of nucleic acid sequences encoding DNA-binding domains that can be used in this capacity include, but are not limited to, sequences encoding a leucine zipper, the bZIP domain, the zinc-finger, the homeobox, the basic helix-loop-helix domain or others. It will be appreciated by those skilled in the art that the selected DNA-binding domain must be capable of interacting with the promoter operatively linked to the one or more genes of interest for which control of expression is desired. Still further, it will be appreciated by those skilled in the art that sequences encoding the acidic domain of RF2a, for example, and at least one DNA-binding domain may be tethered directly to one another, or, alternatively, may be connected indirectly through intervening sequences, e.g., spacers, other polypeptide-encoding sequences, etc.
sequence that is at least substantially similar to SEQ ID NO:6 and is capable of modulating the expression level of the nucleic acid sequence operatively linked to the first promoter through means other than direct DNA binding or interaction. As described above, such transcription factors may be used to regulate the expression level of one or more genes of interest by, for example, affecting the formation of the transcription initiation complex or recruiting other regulatory proteins to the transcription initiation region.

Any promoter, or promoter fragment, capable of driving gene expression in plant cells may be operatively linked to sequences encoding the novel transcription factors of the present invention. The promoter selected for any given embodiment may confer constitutive expression in the transformed plant cell, or, alternatively, inducible expression. Furthermore, as described above with respect to other embodiments, certain promoters are known to be active in particular cell-types, in certain tissues and/or under certain abiotic conditions. Thus, the cell-type and/or conditions in which gene expression is desired will impact the identity of the promoter selected to drive expression of sequences encoding the novel transcription factors of the present invention.

Thus, in one embodiment, the nucleic acid sequences encoding the novel transcription factors of the present invention may be placed under the control of chemically-inducible promoters. In this embodiment, the gene expression system may be activated at will, which provides control over the effect that the encoded novel transcription factors may have on the expression of promoter sequences (and genes operatively linked to such sequences) that comprise the element to which the expressed DNA-binding domain, for example, is capable of interacting. As described above with respect to other embodiments, several inducible promoters well-known in the art could be used in this capacity to drive the expression of novel transcription factors.

It will be appreciated by those skilled in the art that the first promoter, which is operatively linked to at least one gene of interest, and the second promoter, which is operatively linked to a sequence encoding a novel transcription factor of the present invention, may exist in a single gene expression cassette (or vector), or, alternatively, in separate cassettes (or vectors). Still further, those skilled in the art will appreciate that a single promoter, or, alternatively, a plurality of promoters, each operatively linked to at least one gene of interest, may contain a sequence and/or element capable of interacting with the encoded novel transcription factors. Moreover, the plurality of promoters can be substantially similar in sequence, or, alternatively, may comprise significantly different promoters.

Methods of Regulating Gene Expression

The invention further provides methods of regulating the expression level of at least one gene of interest, which comprise transforming a plant cell with at least one plant gene expression cassette of the present invention. Still further, the invention provides methods of regulating the expression level of at least one gene of interest, which comprise (a) constructing at least one plant gene expression cassette of the present invention; (b) transforming a plant cell with the plant gene expression cassette of the present invention; and (c) regenerating whole plants from the transformed plant cell. In a related embodiment, for plants, plant tissues or plant cells that have been transformed with a gene expression cassette comprising a chemically-inducible promoter operatively linked to a nucleic acid sequence that encodes a polypeptide comprising an amino acid sequence that is at least substantially similar to SEQ ID NO:4, 5 or 6, the method further comprises contacting the transformed plants, plant tissues or plant cells, directly or indirectly, with an activating amount of the expression-inducing chemical.

In other embodiments, plant gene expression cassettes comprising two or more chimeric promoters of the present invention, which are operatively linked to one or more genes of interest, can be transformed into a plant cell in connection with, for example, RF2a- and/or RF2b-encoding sequences to achieve a "cascade-type" system. In such embodiments, the expression of RF2a and/or RF2b activates and/or enhances the expression of the nucleic acid sequences to which the two or more chimeric promoters are operatively linked. Still further, the nucleic acid sequences encoding the RF2a and/or RF2b proteins, for example, may be operably linked to chemically-inducible promoters. In such case, upon contacting plants, plant tissues or plant cells, which have been transformed with such sequences, with the expression-inducing chemical, the RF2a and/or RF2b transcription factors are produced. The RF2a and/or RF2b transcription factors subsequently interact with the Box II elements (and/or operational derivatives of Box II) to regulate gene expression. This interaction, of course, results in the synchronized activation and/or enhancement of expression of all Box II-dependent genes (or all genes operatively linked to promoters containing operational derivatives of Box II).

In other embodiments, plant gene expression cassettes comprising at least one chimeric promoter of the present invention, which is operatively linked to one or more genes of interest, can be transformed into plant cells in connection with RF2a- and/or RF2b-encoding sequences to "turn-off" or modulate the expression of one or more unrelated endogenous and/or exogenous genes. The foregoing expression cassettes can be used, for example, to activate and/or enhance the expression of the one or more genes of interest that encode molecules that selectively hybridize to specific target nucleic acid sequences, e.g., endogenous and/or exogenous genes. The hybridization of an oligomeric compound, for example, with its target nucleic acid sequence can have the effect of interfering with the normal function of the target sequence (this effect is generally referred to as "antisense"). The functions of DNA that can be affected in this embodiment, for example, include replication and transcription. The functions of RNA that can be affected include all vital functions such as translocation of the RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be imparted or facilitated by the RNA. The effect of such interference with target nucleic acid function provides the ability to modulate the expression of particular gene products.

The transformation of plant cells in accordance with the present invention may be carried out in essentially any of the various ways known to those skilled in the art of plant molecular biology. That is, the method employed for transformation of target plant cells is not relevant to the present invention and any method suitable for the target plant cell-type may be utilized. As used herein, the term "transformation" refers to the alteration of the genotype of a host plant, including within plant cells, embryos and tissues, by the introduction of exogenous or endogenous nucleic acid sequences. Further, the terms "transfection" and "transformation," as used herein, may be used interchangeably, wherein the meaning accorded to both terms is, generally, as described above with respect to "transformation."
ous embodiments of the present invention may be applied to any plant, plant tissue, seed or plant cell for which transformation techniques are, or become, available.

In general, to commence a transformation process in accordance with the present invention, it is first necessary to construct a suitable vector and properly introduce the vector into a plant cell. The details of the construction of vectors utilized herein are known to those skilled in the art of plant molecular biology. As described above, one or more plant gene expression cassettes may be constructed to practice the methods, and to generate the plants, plant tissues and plant cells of the present invention. In practice, the construct or constructs comprising the expression cassettes of the present invention will be inserted into a plant cell by transformation.

For example, constructs that include chimeric promoters of the present invention, which comprise the Box II element and/or its operational derivatives, can be introduced into plant cells using Ti plasmids, root-inducing (Ri) plasmids, and plant virus vectors. In the first instance, for example, the nucleic acid sequences of the present invention can be introduced into plant cells through Agrobacterium-mediated transformation. Methods involving the use of *Agrobacterium*-mediated transformation include, but are not limited to: 1) co-cultivation of *Agrobacterium* with cultured isolated protoplasts; 2) transforming (or infecting) plant cells or tissues with transformed *Agrobacterium* (as described herein); or 3) transformation of seeds, explants, apices or morristems with *Agrobacterium*. Under appropriate conditions known in the art, the transformed plant cells may be grown to form shoots, roots, and develop further into plants.

In some cases, it may be preferred to introduce the nucleic acid sequences of the present invention into plant cells utilizing *Agrobacterium tumefaciens* containing the Ti plasmid. When using an *A. tumefaciens* culture as a transformation vehicle, it is most advantageous to use a non-oncogenic strain of the *Agrobacterium* as the vector so that normal non-oncogenic differentiation of the transformed tissues is possible. It is also preferred that the *Agrobacterium* harbor a binary Ti plasmid system. Such a binary system comprises 1) a first Ti plasmid having a virulence region essential for the introduction of transfer DNA (T-DNA) into plants and 2) a chimeric plasmid. The chimeric plasmid contains at least one border region of the T-DNA region of a wild-type Ti plasmid flanking the nucleic acid sequence to be transferred. Binary Ti plasmid systems have been shown to be effective in transforming plant cells.

Alternatively, the nucleic acid sequences of the present invention can be introduced into plant cells using mechanical or chemical means. For example, nucleic acid sequences can be mechanically transferred by direct microinjection into plant cells utilizing micropipettes. Still further, the nucleic acid sequences may be transferred into plant cells using polyethylene glycol, which is capable of forming a precipitation complex with nucleic acid sequences that is taken up by target plant cells.

The nucleic acid sequences of the present invention can also be introduced into plant cells by electroporation. In this technique, plant protoplasts, for example, are electroporated in the presence of vectors or nucleic acid sequences to be transformed into the protoplasts. Electrical impulses of high field strength reversibly permeabilize plant membranes allowing the introduction of nucleic acids. Electroporated plant protoplasts reform the cell wall, divide and form a plant callus. Selection of the transformed plant cells with the transformed gene can be accomplished using, for example, phenotypic markers.

Another well-known method for introducing nucleic acid sequences of the present invention into plant cells is high velocity BIOLISTICO® penetration by small particles with the nucleic acid sequences to be introduced contained either within the matrix of small beads or particles, or on the surface thereof. See, for example, U.S. Pat. Nos. 5,932,479 and 5,693,507.

Additionally, DNA viruses may be used as vectors for introducing heterologous nucleic acid sequences into plant cells. See, for example, U.S. Pat. No. 4,407,956. Non-limiting examples of such DNA viruses include the Cauliflower mosaic virus ("CaMV") and the Geminivirus. The CaMV viral DNA genome, for example, may be inserted into a parent bacterial plasmid creating a recombinant DNA molecule which can be propagated in bacteria. After cloning, the recombinant plasmid may be re-cloned and further modified by introduction of the desired nucleic acid sequence of the present invention. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the target plant cells.

In any of the foregoing methods of transformation, a selectable marker may, optionally, be associated with constructs comprising nucleic acid sequences of the present invention. As used herein, “marker” refers to a gene that encodes a protein that confers a particular trait or a phenotype that permits the selection of, or the screening for, a plant or plant cell containing the marker. In some cases, the marker gene may encode a protein that confers antibiotic resistance to transformed plant cells, whereby the appropriate antibiotic can be used to select for transformed plant cells among cells that are not transformed. Examples of suitable selectable markers include adenosine deaminase, dihydrofolate reductase, hygromycin-B-phosphotransferase, thymidine kinase, xanthine-guanine phosphoribosyltransferase and amino- glycoside 3′-O-phosphotransferase II (kanamycin, neomycin and G418 resistance). Other suitable markers will be known to those of skill in the art. For example, screenable markers, such as the uidA gene, which encodes β-glucuronidase ("GUS"), luciferase or the gene encoding the green fluorescent protein ("GFP"), may also be used.

Plants and Plant Parts

Still further, the invention provides plant cells, plant embryos, plant tissues, whole plants and seeds that have been transformed with at least one plant gene expression cassette of the present invention. Methods of regenerating whole plants from transformed plant cells, embryos and tissues are well-known to those skilled in the art.

The following Examples are included to demonstrate certain embodiments of the invention. It should be appreciated by those skilled in the art that the techniques disclosed in the Examples that follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus, can be considered to constitute preferred modes for its practice. However, those of ordinary skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

EXAMPLES

Example I

Analysis of the Box II Sequence in BY-2 Protoplasts

Until now, the behavior of the RTBV promoter in the tobacco BY-2 cell line and the contribution of its various cis elements to the promoter activity in BY-2 cells was unknown.
To compare the relative activity of the RTBV promoter with other constitutive promoters, tobacco BY-2 protoplasts were transfected with gene constructs in which the uidA open reading frame (“ORF”) was driven by the enhanced cauliflower mosaic virus ("CaMV") 35S promoter, the enhanced CaMV promoter with a 5' deletion at position -68, and -32, respectively. In FIG. 2, “pE” represents the entire E fragment; “p-100” represents a truncated E fragment at position -100, which excludes the GATA motif; “p-68” represents a truncated E fragment at position -68, which excludes the GATA motif and ASL box; and “p-32” represents a truncated E fragment at position -32, which excludes the GATA motif, ASL box and Box II element. All of the foregoing fragments included the TATA box and Box I element of the RTBV promoter.

The mutated promoters were then fused with the uidA gene. The GUS expression levels of the derived plasmids were determined in BY-2 protoplasts. The data in FIG. 2 show that the Box II element is crucial for promoter function in BY-2 protoplasts. As long as the Box II element was retained, the minimal promoter exhibited an expression level similar to that of the full E fragment. Once the Box II element was removed, however, the promoter activity decreased to less than 20% of the E fragment.

Example 2

The Optimal Position of the Box II Sequence

The optimal spacing of the Box II element, relative to the TATA box, for maximal effect on expression of the host promoter was determined. The original Box II sequence was removed and inserted in locations other than its native location (see Table 1:).

<table>
<thead>
<tr>
<th>Promoters</th>
<th>Primer Sets</th>
<th>Respective SEQ ID Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-Box II+11</td>
<td>R/mtEBoxII(-0.5 h)-3'</td>
<td>SEQ ID NO: 73/14</td>
</tr>
<tr>
<td>E-Box II-9</td>
<td>R/mtEBoxII(-0.5 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/15</td>
</tr>
<tr>
<td>E-Box II-17</td>
<td>R/mtEBoxII(-1.0 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/17</td>
</tr>
<tr>
<td>E-Box II-58</td>
<td>R/mtEBoxII(-2.5 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/20</td>
</tr>
<tr>
<td>E-Box II-63</td>
<td>R/mtEBoxII(-4.5 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/22</td>
</tr>
<tr>
<td>E-Box II-74</td>
<td>R/mtEBoxII(-6.0 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/24</td>
</tr>
<tr>
<td>E-Box II-79</td>
<td>R/mtEBoxII(-9.0 h)-5'/GUS3'</td>
<td>SEQ ID NO: 73/26</td>
</tr>
<tr>
<td>E-Box II-100</td>
<td>R/mtEBoxII(-111 nt)-3'</td>
<td>SEQ ID NO: 73/28</td>
</tr>
<tr>
<td>E-Box II-116</td>
<td>R/mtEBoxII(-111 nt)-5'/GUS3'</td>
<td>SEQ ID NO: 73/29</td>
</tr>
<tr>
<td>E-Box II-147</td>
<td>R/mtEBoxII(+29 nt)-3'</td>
<td>SEQ ID NO: 73/32</td>
</tr>
<tr>
<td>E-Box II-164</td>
<td>R/mtEBoxII(+60 h)-3'</td>
<td>SEQ ID NO: 73/34</td>
</tr>
</tbody>
</table>

The PCR products generated in the first reaction were used as templates for the second PCR reaction using R/GUS3' as primers (SEQ ID NO:73/13). The PCR products from the second reactions were purified and cloned into the pE:GUS vector to replace the E promoter through the restriction sites HindIII/NcoI. The resultant plasmids were named pEBoxII-11:GUS, pEBoxII-9:GUS, pEBoxII-17:GUS, pEBoxII-58:GUS, pEBoxII-63:GUS, pEBoxII-74:GUS, pEBoxII-116:GUS, pEBoxII-147:GUS and pEBoxII-164:GUS, respectively.

BY-2 protoplasts were transfected with the foregoing constructs and tested for GUS activity. Protein samples from such protoplasts were prepared using protein extraction buffer (Jefferson et al., 1987) and quantified using the DC protein assay kit (Bio-Rad Laboratories, Hercules, Calif.). Quantitative analysis of GUS activity was performed as described by Jefferson et al. (1987) using the substrate 4-methylumbelliferyl-b-D-glucuronide (“MUG”) with the Spectra Max Gemini instrument ( Molecular Devices Corp., Sunnyvale, Calif.).

As shown in FIG. 4, the E promoter activity dramatically decreased when the Box II element was removed (see construct pEBoxII). Additionally, when the Box II element was removed and inserted in locations other than its native location, the mutated promoters exhibited similar activity as pE-
BoxII, which indicates that the position of the Box II element in its native promoter is important.

Example 3

Effect of RF2a and RF2b on Expression of a Reporter Gene

As shown in FIG. 1, the E fragment of the RTBV promoter showed strong activity in tobacco BY-2 protoplasts, presumably because the BY-2 cell line contains RF2a- and RF2b-like transcription factors (Fukazawa et al., 2000). Nevertheless, the E fragment may be further stimulated by co-transfecting the BY-2 protoplasts with constructs that encode the RF2a and/or RF2b transcription factors. In this Example, the E fragment was activated in tobacco BY-2 protoplasts by co-transfection of: E:GUS with CaMV::RF2b, CaMV 35S::RF2a and CaMV 35S::RF2a/CaMV::RF2b using methods well-known in the art. As shown in FIG. 5, the E fragment can be activated above a strong background expression by over-expression of RF2a and/or RF2b.

Example 4

Mutants of Box II

To investigate the possibility of reducing the background expression level in BY-2 protoplasts, the activities of certain Box II mutants (or “operational derivatives”), as shown in FIG. 6, were tested in BY-2 protoplasts. Specifically, the operational derivatives Box IIm1 (SEQ ID NO:2) and Box IIm2 (SEQ ID NO:3) were tested in the context of the E fragment (SEQ ID NO:3 and SEQ ID NO:64, respectively). The mutated promoters were ligated with the uidA coding sequence to create fusion genes, which are referred to herein as pE(Box IIm1)::GUS and pE(Box IIm2)::GUS. BY-2 protoplasts were then transfected with the pE(Box IIm1)::GUS and pE(Box IIm2)::GUS constructs. When the Box IIm1 and Box IIm2 elements were used, the GUS activity relative to the wild type Box II element (pE::GUS) dropped significantly (FIG. 7).

In the case of the chimeric sequence pE(Box IIm2)::GUS, there was little, if any, GUS activity above that of non-transfected BY-2 cells (FIG. 7). These data suggest that there are endogenous transcription factors in BY-2 cells that can interact with the wild type Box II element, which results in expression of pE::GUS in protoplasts (FIGS. 1, 2 and 7). When operational derivatives of the Box II element were used, e.g., Box IIm1 and Box IIm2, the promoter activity was significantly abolished.

Example 5

Binding Affinities of RF2a and RF2b

A previous report showed that RF2a binds to the Box II element and its mutants with different affinities (Yin et al., 1997). To compare the binding affinities of RF2a and RF2b with the Box II element and its mutants, real time Surface Plasmon Resonance (“SPR”) measurements were conducted using a BLAcore 2000 instrument. The binding affinities of RF2a and RF2b to these DNA targets were measured on chips on which biotin labeled-Box II, -Box IIm1 and -Box IIm2 elements were immobilized.

The association and de-association constants were determined using BLAevaluation 3.1 software—using 1:1 binding with a mass transfer model. The results are presented in FIG. 9 for the Box IIm1 element, and FIG. 10 for the Box IIm2 element.

When the foregoing constructs were co-transfected with CaMV 35S::RF2a, the promoter activity of E(Box IIm1) and E(Box IIm2) increased 5 to 7 fold (FIGS. 9 and 10). Furthermore, there was no apparent difference between the activation of promoters containing the Box IIm1 and Box IIm2 elements. Different results were observed when the foregoing constructs were co-transfected with CsVMV::RF2b. The E(Box IIm2) promoter was activated about 5.8 fold by RF2b, while the E(Box IIm1) promoter was activated about 7 fold (FIGS. 9 and 10). Importantly, this Example shows that the activity of the E(Box IIm1) promoter with RF2b was as high as the expression of the E wild type promoter in BY-2 protoplasts.

The foregoing data related to the interactions between the RF2a and/or RF2b transcription factors and the target chimeric promoters suggest that such promoters may, optionally, be designed to comprise the Box IIm1 and/or Box IIm2 elements of the chimeric promoters, using compositions and methods described herein.

Example 6

Use of Box II and RF2a to Control Expression of Novel Chimeric Promoters

The following shows that the Box II element can be transferred to unrelated promoters, or promoter fragments, in a position dependent manner to control heterologous gene expression. To show that the Box II element is portable, Box II was fused with different lengths of the CaMV 35S promoter (FIG. 11). Specifically, the Box II element was fused to the 5’ end of the chimeric promoters with a space of 7 nucleotides from the TATA box plus 1, 1.5, 3.0, 3.5, 5.0 and 5.5 “turns” of DNA helix” (one turn = 10.4 base pairs). The chimeric promoters were then inserted into a cassette with the uidA coding sequencing and Nos terminator.

Construction of Plasmids Comprising the Box II Element and Different Lengths of the CaMV 35S Promoter—To construct Box II and CaMV 35S chimeric promoters, the Box II element was introduced into CaMV 35S promoter sequences of different lengths through PCR reactions using Pfu DNA polymerase (Stratagen Systems, Kirkland, Wis.). Forward
primers for the PCR reactions were designed to have a HindIII restriction site, followed by the Box II sequence, which was followed by part of the 5'CaMV promoter at desired positions (see FIG. 12 and Table 2 for specific primer sequences). The GUS 3' primer (SEQ ID NO:36) was used as reverse primer for all reactions.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Primer Set</th>
<th>5' Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.0hBoxII-48CaMV</td>
<td>-1.0hBoxII-48CaMV/GUS 3' SEQ ID NO:37</td>
</tr>
<tr>
<td>2</td>
<td>-1.5hBoxII-53CaMV</td>
<td>-1.5hBoxII-53CaMV/GUS 3' SEQ ID NO:38</td>
</tr>
<tr>
<td>3</td>
<td>-3.0hBoxII-69CaMV</td>
<td>-3.0hBoxII-69CaMV/GUS 3' SEQ ID NO:39</td>
</tr>
<tr>
<td>4</td>
<td>-3.5hBoxII-74CaMV</td>
<td>-3.5hBoxII-74CaMV/GUS 3' SEQ ID NO:40</td>
</tr>
<tr>
<td>5</td>
<td>-5.0hBoxII-90CaMV</td>
<td>-5.0hBoxII-90CaMV/GUS 3' SEQ ID NO:41</td>
</tr>
<tr>
<td>6</td>
<td>-5.5hBoxII-95CaMV</td>
<td>-5.5hBoxII-95CaMV/GUS 3' SEQ ID NO:42</td>
</tr>
</tbody>
</table>

The p35S:GUS plasmid was used as template in the foregoing reactions. The p35S:GUS plasmid, a NcoI restriction site was present in all PCR products. The PCR products were restricted with HindIII/NcoI and cloned into a pE:GUS vector to replace the E promoter through the same set of restriction sites. The constructed plasmids containing the chimeric promoters labeled 1 through 6 in Table 2 were named p-1.0hBoxII-48CaMV:GUS, p-1.5hBoxII-53CaMV:GUS, p-3.0hBoxII-69CaMV:GUS, p-3.5hBoxII-74CaMV:GUS, p-5.0hBoxII-90CaMV:GUS, and p-5.5hBoxII-95CaMV:GUS, respectively. The controls for this set of plasmids were generated using primers sets 1.5h-53CaMV-c (SEQ ID NO:38)/GUS 3' and -3.5h-74CaMV-c (SEQ ID NO:40)/GUS 3' and -5.5h-95CaMV-c (SEQ ID NO:42)/GUS 3', which products were cloned into pE:GUS with HindIII/NcoI to replace the E promoter. The derived plasmids/constructs were named p-53CaMV:GUS, p-74CaMV:GUS and p-95CaMV:GUS, respectively.

The foregoing constructs were then transfected into non-transgenic BY-2 protoplasts, and the resulting GUS activity determined. The data in FIG. 13 show that when the Box II element is fused with the CaMV 35S promoter with a space of 7 nucleotides from the TATA box plus one or more full turns of DNA helix (p1.0hBoxII-48CaMV and p5.0hBoxII-90CaMV), gene expression was dramatically stimulated, compared to reporter genes that lack Box II (p-48CaMV and p-90CaMV, respectively). The p1.0hBoxII-48CaMV and p5.0hBoxII-90CaMV constructs yielded 6.9 and 4.7 fold increases in expression level above p-48CaMV and p-90CaMV, respectively, which did not contain the Box II element.

In contrast, when the Box II element was fused to fragments of the CaMV 35S promoter at positions of 7 nucleotides from the TATA box plus multiples of 0.5 turns of DNA helix (p-1.5hBoxII-53CaMV versus p-53CaMV; p-3.5hBoxII-74CaMV versus p-74CaMV; and p-5.5hBoxII-95CaMV versus p-95CaMV), there was much less or no stimulation of gene expression. Of course, these results indicate that the Box II element, and its operational derivatives, can regulate the expression of unrelated promoters, e.g., the CaMV 35S chimeric promoters described above, as it can in the RTBV promoter. Furthermore, the data indicate that the effect of the Box II element, and its operational derivatives, is, preferably, imparted in a position and/or orientation dependent manner—as described above.

Example 7

Validation of Transient Assay Data From Tobacco
BY-2 Protoplasts in Transgenic Arabidopsis

To evaluate the data presented above from the transient assays, binary vectors were built and transformed into Arabidopsis plants through Agrobacterium-mediated transformation. The set of binary vectors that were transformed into plants were constructs with different deletions of the RTBV promoter, which comprised the various portions of the E fragment described in Example 1; the E fragment containing either the Box IIm1 or Box IIm2 element; or the wild type E fragment (See FIG. 15A). In the transformed Arabidopsis plants, the activity of the chimeric promoters, which comprised the Box IIm1 or Box IIm2 element, was near the basal level of expression, which was observed for the construct p-E(32):GUS, in which all the cis elements of RTBV promoter, which comprised the various portions of the E fragment described in Example 1; the E fragment containing either the Box IIm1 or Box IIm2 element; or the wild type E fragment (See FIG. 15A). In the transformed Arabidopsis plants, the activity of the chimeric promoters, which comprised the Box IIm1 or Box IIm2 element, was near the basal level of expression, which was observed for the construct p-E(32):GUS, in which all the cis elements of RTBV promoter up-stream of the TATA box were removed (FIG. 15B). Thus, the foregoing data agree with the data generated in the transient protoplasts analysis described in Example 5, wherein the chimeric promoters of the present invention may, for example, be designed to comprise the operational derivatives Box IIm1 and/or Box IIm2 to create a near “zero” background expression level in the absence of RF2a and/or RF2b.

Example 8

RF2a Mutants with Deletions of Functional Domains

It has been shown that the bZIP protein RF2a enhances transcription in vivo and in vitro. It is further known that the RF2a transcription factor comprises a proline-, acidic- and glutamine-rich domain. To analyze the function of each domain, mutants of RF2a were created by removing one or more of the foregoing putative domains as shown in FIG. 16(A). More particularly, mutants of RF2a were created in which the proline-rich domain was removed (RF2a-ΔP); the
glutamine-rich domain was removed (RF2a-DQP); the proline-rich and acidic domains were removed (RF2a-DPPA); the proline- and glutamine-rich domains were removed (RF2a-DPMQ); and the glutamine-rich, proline-rich and acidic domains were removed (RF2a-3A). The coding sequence for each mutant was cloned into the bacterial expression vector pET28a, in which a His, tag was placed at the N-terminus of the fusion protein. The derived plasmids were named pET-RF2a (encoding full-length RF2a), pET-RF2a-DP, pET-RF2a-DQ, pET-RF2a-DPA, pET-RF2a-DPMQ, and pET-RF2a-3A.

Plasmid Construction for Protein Purification—The sequences encoding the foregoing mutants of RF2a were created through PCR amplification. A NdeI restriction site was added to the 5' end of all primers, and the ATG in the restriction site was in frame with the His, tag in vector pET28a (Invitrogen Corp., Carlsbad, Calif.) and served as the transcription start codon for the plasmids described in Example 9. A BamHI site was added to all the 3' primers with a stop codon in front of the restriction site. The primers used for amplification of the various fragments of RF2a are listed in Table 3 below:

| RF2a 5'  | SEQ ID NO: 46 |
| RF2a-DP 5' | SEQ ID NO: 47 |
| RF2a-DPA 5' | SEQ ID NO: 48 |
| RF2a 3' | SEQ ID NO: 49 |
| RF2a-DQ 3' | SEQ ID NO: 50 |

From a complete RF2a-encoding sequence (SEQ ID NO: 7), the ΔP fragment was amplified using primers RF2a-ΔP 5' and RF2a 3'; ΔQ was amplified using primers RF2a 5' and RF2a-AQ 3'; ΔPA was amplified using primers RF2a-DPA 5' and RF2a 3'; and ΔPAQ was amplified using primers RF2a-ΔP 5' and RF2a-AQ 3'. The construction of pET-RF2a-3A and pET-RF2a were described by Petruccelli et al. (2001). All of the fragments were restricted with NdeI and BamHI and were cloned into pET28a through the same set of restriction sites. All of the mutations were verified by DNA sequence analysis. The derived plasmids were designated pET-RF2a-ΔP, pET-RF2a-AQ, pET-RF2a-DPA, pET-RF2a-ΔPAQ, and pET-RF2a (encoding full-length RF2a), and pET-RF2a-3A.

Protein Purification—The pET28a-derived plasmids were transformed into Escherichia coli strain BL21 (DE3)pLysS for protein expression. Protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside at room temperature for 3 hours after the cell density reached A600 of —0.6. The His-tagged proteins were purified according to procedures provided by Novagen, Inc. (Madison, Wis.) under non-denaturing conditions. The purified recombinant proteins were dialyzed in lx phosphate-buffered saline with 20% glycerol to remove imidazole and stored at —70° C.

The purified mutant proteins were then analyzed by SDS-PAGE to confirm that each protein was its expected size (FIG. 16(C)). The data presented in FIG. 16(C) demonstrate that proteins ΔP (SEQ ID NO:68), ΔQ (SEQ ID NO:69), ΔPA (SEQ ID NO:70), ΔPAQ (SEQ ID NO:71), and 3A (SEQ ID NO:72) of RF2a bind to the Box II element.

Example 9

Contribution of RF2a Domains to Activity

The relative activity of RF2a and the RF2a mutants (described in Example 8) was then measured. First, a chimeric promoter was developed with a single copy of the Box II element fused to the 5' end of a minimal CaMV 35S promoter comprising nucleotides —48 to +8 (SEQ ID NO:66). The chimera promoter was ligated to the uidA coding sequence to create the reporter pBI1-48Ca:GUS.

To analyze the function of the several domains of RF2a, effectors were created by inserting coding sequences of RF2a or the mutants of RF2a (described in Example 8) downstream of the enhanced CaMV 35S promoter in the pMON999 vector (a gift from Monsanto Company, St. Louis, Mo.). The resultant constructs, p35S::RF2a, p35S::RF2a-AP, p35S::RF2a-ΔQ, p35S::RF2a-DPA, p35S::RF2a-AQ and p35S::RF2a-3Δ, were co-transfected into BY-2 protoplasts with pBI1-48Ca::GUS (FIG. 17(A)). Plasmid pCat-GFP, in which the GFP gene was driven by CaMV 35S promoter, was co-introduced to serve as an internal control. The following describes, in greater detail, the construction of these vectors and the methods employed in transfecting the same into BY-2 protoplasts.

Plasmids for protoplast transfection—The coding sequences for mutants of RF2a were released from pET28a-derived plasmids and cloned into the plant expression vector pMON999 (a gift from Monsanto Company, St. Louis, Mo.) to place each gene downstream of an enhanced CaMV 35S promoter, followed by a nopaline synthase terminator sequence. The resulting effector constructs were named p35S::RF2a, p35S::RF2a-AP, p35S::RF2a-ΔQ, p35S::RF2a-DPA, p35S::RF2a-AQ and p35S::RF2a-3Δ. The reporter gene construct, pBI1-48Ca::GUS, was built using PCR to introduce the Box II element into a minimal CaMV 35S promoter comprising nucleotides —48 to +8 with primers BII-48Ca 5' (SEQ ID NO:53) (which contained the Box II element) and GUS 3' (SEQ ID NO:36) using a p35S::GUS plasmid as template. The PCR product was restricted with HindIII and Ncol, and the resulting fragment was inserted into p35S::GUS to replace the original 35S promoter.

Transfection of tobacco BY-2 protoplasts—The protoplasts were isolated from tobacco cell line BY-2 as described by Watanabe et al. (1987). Approximately one million protoplasts were transfected by electroporation with 20 µg of effector construct DNA, 15 µg of herring sperm DNA, 2.5 µg of reporter gene construct DNA, and 15 µg of pCaI-GFP DNA. In samples with reporter gene alone, the total amount of DNA was adjusted by adding 20 µg of herring sperm carrier DNA. The electroporation parameters used were 300 V and 250 microfarads with the Bio-Rad electroporation system (Bio-Rad Laboratories, Hercules, Calif.). Protoplast samples were cultured in Murashige and Skoog medium with 0.4 M mannitol, pH 5.8, at 28° C. The protoplasts were collected 24 hours after electroporation.

As shown in FIG. 17(B), the transactivation function of RF2a was not decreased by removing either the proline-rich (35S::RF2a-AP) or glutamine-rich (35S::RF2a-AQ) domains or both of the domains (35S::RF2a-ΔPAQ). In fact, the activation function of each of these mutants was greater than that of full-length RF2a. RF2a-AP was significantly different from RF2a at the P0.05 level, whereas RF2a-DQ and RF2a-
ΔPAQ) were significantly different from RF2a at the P<0.01 level (Student’s t test). Also, the difference between the activity of RF2a-ΔP and RF2a-ΔQ was significant at the P<0.05 level, and there was no difference between RF2a-ΔQ and RF2a-ΔPAQ.

The data suggest that the proline-rich and glutamine-rich domains do not contribute in a positive way to the activation function of RF2a. In contrast, the activity dropped to near basal level when the acidic domain was removed (RF2a-ΔPA and RF2a-3A) (FIG. 17(B)). These results suggest that the acidic domain is responsible for the activation of gene expression by RF2a.

Example 10
Functions of RF2a Domains in Fusion Proteins

To determine whether domains of RF2a can serve as independent modules to regulate transcription, the various putative functional domains were fused with the synthetic 2C7 protein (SEQ ID NO:75), a synthetic zinc finger DNA-binding domain that specifically binds to the 2C7 DNA-binding site (SEQ ID NO:74), to create various “effector” constructs (FIG. 18(A)). The various RF2a domains were placed either at the N-terminus or the C-terminus of the 2C7 DNA-binding domain (“DBD”). The “reporter” construct (PC7er2:GUS) carried the uidA coding sequence located downstream of a chimeric promoter comprising 6x2C7-binding sites ligated with the minimal promoter of erbl-2 (“er2”), p35S:2C7 encoded the 2C7 protein without an activation domain and served as a control (FIG. 18(A)).

To create effectors with RF2a domains fused to the N-terminus of 2C7 DBD, coding sequences for the acidic domain (A) (SEQ ID NO:9), proline-rich domain (P) (SEQ ID NO:54) and glutamine-rich domain (Q) (SEQ ID NO:55) were amplified using primer pairs A-2C7 5'/A-2C7 3', P-2C7 5'/P-2C7 3', and Q-2C7 5'/Q-2C7 3', respectively, with pET-RF2a as template. BglII and BamHI restriction sites were introduced into the 5' and 3' primers, respectively. The particular sequences of the foregoing primers are referenced in Table 4 below:

| A-2C7 5'  | A-2C7 3'  | SEID NO: 56 | SEID NO: 57 |
| P-2C7 5'  | P-2C7 3'  | SEID NO: 58 | SEID NO: 59 |
| Q-2C7 5'  | Q-2C7 3'  | SEID NO: 60 | SEID NO: 61 |

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

The products created by the foregoing PCR reactions were restricted with the BglII and BamHI and cloned into pMON999-derived plasmids using NotI (blunted) and cloned into the binary vector pGA-E::GUS (Petruccelli et al., 2001) using the blunt HindIII site (downstream of a CaMV 35S promoter sequence). The final plasmids were named pGA-E::GUS/P-35S::ΔP, pGA-E::GUS/P-35S::ΔQ, pGA-E::GUS/P-35S::ΔPA, and pGA-E::GUS/P-35S::ΔPAQ. A plasmid encoding full-length RF2a, p35S::RF2a, was also constructed using the methods described herein.

Tobacco transformation—the plants described above were introduced into Agrobacterium tumefaciens (strain LBA4404) and used for tobacco transformation. Leaf discs from Nicotiana tabacum cv. Xanthi NN were transformed with the various plasmids following the protocol of Horsch et al. (1988). At least 15 independent transgenic lines were produced with each gene construct. Transgenic plants were self-fertilized, and T1 seeds were collected. The T1 seeds were germinated on Murashige and Skoog medium (Murashige and Skoog, 1962) with kanamycin (100 mg/L) selection, and Kan' seedlings were grown in a greenhouse for observation.

Following PCR analysis, transgenic lines expressing each mutant were observed for phenotypic changes. T1 generation plants with 35S::RF2a, 35S::RF2a-ΔP, 35S::RF2a-ΔQ, and 35S::RF2a-ΔPAQ did not exhibit abnormal phenotypes (FIG. 19(A)). However, 11 of 15 independent transgenic lines with 35S::RF2a-ΔPA exhibited mild to severe stunting with curved leaves and substantial delay in flowering times (FIGS. 19(A) & (B)). Furthermore, the internodal elongation of transgenic plants was strongly repressed by RF2a-ΔPA (FIG. 19(C), Panel 1). The phenotype caused by 35S::RF2a-ΔPA was similar to, but less severe than, the phenotype caused by 35S::RF2a-3A. Cross-sections of the stem of transgenic plants with either RF2a-ΔPA or RF2a-3A showed that
the xylem of stunted plants was not uniformly lignified and that phloem development was altered. (FIG. 19(C)).

To confirm that the phenotype was related to transgene expression, protein extracts of the transgenic plants were analyzed via a Western blot assay using an antibody against RF2a. Protein samples from the transgenic leaf tissues were extracted in buffer (50 mM NaPO₄, pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1% sodium lauryl sarcosine) and quantified using the DC protein assay kit (Bio-Rad Laboratories, Hercules, Calif.). 40 µg of such protein sample was separated via SDS-PAGE and blotted onto nitrocellulose membrane. The membrane was stained with Ponceau S (Sigma Chemical Company, St. Louis, Mo.) to monitor protein loading prior to immuno-detection. The primary antibody used in the immuno-detection was raised in rabbits against full-length RF2a; the secondary antibody was horseradish peroxidase-conjugated to goat anti-rabbit antibody (Southern Biotechnology Associates, Inc., Birmingham, Ala.). FIG. 20 shows that there is a direct correlation between the abnormal phenotype and the accumulation of RF2a-APAA.

Example 12

Versatility of the Gene Regulation System

As shown herein, the acidic domain of RF2a may be linked to any DNA binding domain to regulate the expression of corresponding promoters. More specifically, the novel transcription factors of the present invention that comprise the acidic domain of RF2a (or substantially similar sequences) and at least one DNA binding domain (“DBD”), may be used to regulate the expression of plant functional promoters that comprise one or more cis elements—wherein such elements are capable of interacting with such DNA binding domain. The interaction between such novel transcription factors through such DNA binding domains with corresponding cis elements, preferably, results in the initiation, or enhancement of, transcription.

Similarly, the chimeric promoters of the present invention have been shown to regulate transcription in the presence of RF2a and/or RF2b. More particularly, the inventors have demonstrated that the Box II element and/or its operational derivatives described herein, the RF2a and RF2b transcription factors (and related bZIP proteins) have been found to interact with certain additional cis elements to impart regulation of transcription. Non-limiting examples of such additional cis elements are summarized in Table 5 below:

<table>
<thead>
<tr>
<th>Common Name of cis Element DNA Sequence</th>
<th>SEQ ID NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC-II</td>
<td>CCACACCCACCCAC</td>
</tr>
<tr>
<td>4CL-1</td>
<td>CTTACCACCCCACACT</td>
</tr>
<tr>
<td>Sh1</td>
<td>TG2ACCTCAACA</td>
</tr>
<tr>
<td>AHA3</td>
<td>AGSTACCCCAT</td>
</tr>
<tr>
<td>Vs-1</td>
<td>TGGAATGGAAGCAGCA</td>
</tr>
</tbody>
</table>

The present invention contemplates that such additional cis elements, including, without limitation, the elements referenced in Table 5, may be used to construct novel chimeric promoters in accordance with the present invention. Of course, such novel chimeric promoters may be used, for example, to regulate gene expression in connection with RF2a and/or RF2b transcription factors—similar to the chimeric promoters shown and described in previous Examples. It will be appreciated by those skilled in the art that chimeric promoters of the present invention, which comprise, for example, one or more of the additional cis elements listed in Table 5 (or substantially similar sequences), may be constructed and used in a manner similar to that described above with respect to chimeric promoters comprising Box II and/or its operational derivatives.

Still further, the present invention contemplates that the acidic domain of RF2a (SEQ ID NO: 6), or substantially similar sequences, may be used with other DNA binding domains in the art. More particularly, the present invention contemplates that the acidic domain of RF2a, or substantially similar sequences, may be linked to any DNA binding domain from a plurality of classes of such domains to create novel transcription factors capable of regulating gene expression (which was demonstrated in Example 10). The invention provides, for example, that the following classes of DNA binding domains may be used in such capacity: (i) basic helix-loop-helix domains (“bHLH”); (ii) DNA binding domains of bZIP proteins; (iii) native or synthetic zinc finger DNA binding domains; and (iv) DNA binding domains of the E2F/DP family of transcription factors. Non-limiting examples of such DNA binding domains for each class are listed in Table 6 below. Additionally, Table 6 references the specific cis element with which such DNA binding domains are known to interact (and affect gene expression).

<table>
<thead>
<tr>
<th>Class of DBD</th>
<th>Representative DBD</th>
<th>Sequence of DBD</th>
<th>cis Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>bHLH</td>
<td>“bHLH/Z domain of H. Sapiens”</td>
<td>SEQ ID NO: 84</td>
<td>SEQ ID NO: 85</td>
</tr>
<tr>
<td>bZIP</td>
<td>“Jun” - from H. Sapiens</td>
<td>SEQ ID NO: 86</td>
<td>SEQ ID NO: 87</td>
</tr>
<tr>
<td>Zinc-Finger</td>
<td>“C2H2”</td>
<td>SEQ ID NO: 88</td>
<td>SEQ ID NO: 89</td>
</tr>
<tr>
<td>E2F/DP</td>
<td>“E2F4”</td>
<td>SEQ ID NO: 90</td>
<td>SEQ ID NO: 91</td>
</tr>
</tbody>
</table>

Many of the DNA binding domains and cognate cis elements referenced in Table 6 are described in the literature (among others). See, for example, Ferre-D’Amare, R. et al. (1994); Toledo-Ortiz, G. et al. (2003); Jakoby, M. et al.
expression of appropriately designed chimeric promoters. Transformation was conducted using the well-known dipping domains referenced in Table 6 may be used to regulate the RF2a sequence in pSL-5G35Sm:RF2a through HindIII (blunted)/BamHI restriction. The excised fragment was then made blunt using Klenow treatment in the presence of dNTPs.

The excised GUS- and RF2a-encoding fragment was subsequently cloned into the binary vector pCa-5GRbm:DsRed

expression of corresponding promoters (which contain one or more cis elements that may interact with such DNA binding domain to regulate transcription).

Example 13

Use of Inducible Promoters with the Gene Expression System

The gene expression system of the present invention (including the chimeric promoters, gene expression cassettes, and novel transcription factors described above) may, optionally, be used in connection with a plurality of inducible promoters. In certain preferred embodiments, the present invention is used in connection with chemically-inducible promoters. The following provides a non-limiting example of such embodiments of the present invention.

Plasmid construction—The coding sequence of RF2a was released from a cassette comprising the CaMV 35S promoter operably linked to a RF2a-encoding sequence, p35S::RF2a, through EcoRI/BamHI restriction. The excised DNA fragment was made blunt by Klenow treatment in the presence of dNTPs. The resulting fragment was then cloned into plasmid RH5 (Rohm & Haas, Philadelphia, Pa.) to replace a luciferase coding sequence. The R13 plasmid originally comprised a luciferase coding sequence downstream of a chimeric promoter that included five (5) repeats of the Gal4 DNA binding site and the CaMV 35S minimal promoter. The chimeric promoter comprising the Gal4 DNA binding sites and the CaMV 35S minimal promoter is set forth in SEQ ID NO. 92. The RF2a-encoding sequence, together with the chimeric promoter, were released from the resultant plasmid through SalI/BamHI restriction and inserted into vector pSL301, which was previously linearized using the same set of restriction enzymes. The chimeric promoter and RF2a-encoding sequence are referred to herein as the “5G35Sm::RF2a” sequence. The resulting plasmid is identified herein as “pSL-5G35Sm::RF2a.”

A DNA fragment containing the uidA gene operably linked to the E fragment of the RTBV promoter was released from pE-GUS through HindIII (blunted)/BamHI restriction. The excised fragment was then inserted upstream of the 5G35Sm:RF2a sequence in pSL-5G35Sm:RF2a through Ndel (blunted)/BamHI restriction sites (pSL-E-GUS/5G35Sm:RF2a). The sequences encoding E:GUS, together with 5G35Sm:RF2a, were released from pSL-E-GUS/5G35Sm:RF2a through HindIII restriction. The excised fragment was then made blunt using Klenow treatment in the presence of dNTPs.

The novel transcription factors of the present invention, which comprise, for example, one or more DNA binding domains referenced in Table 6 may be used to regulate the expression of appropriately designed chimeric promoters. More particularly, such transcription factors may be used to regulate the expression of chimeric promoters that comprise, for example, the corresponding cis element referenced in Table 6 (or substantially similar sequences). In light of the foregoing, it should be appreciated that the acidic domain of RF2a (or substantially similar sequences) may be linked to any DNA binding domain known in the art (including, without limitation, the domains listed in Table 6) to regulate the expression of corresponding promoters (which contain one or more cis elements that may interact with such DNA binding domain to regulate transcription).

Analysis of GUS activity—The relative GUS activity in the transgenic Arabidopsis plants described herein was measured by histochemical analysis. In each of sixty-seven (67) T1 transgenic Arabidopsis plants carrying E:GUS, 5G35Sm:RF2a and Cs:VGE, the 1:8000 dilution of Intrepid® 2F (Dow AgroSciences, Indianapolis, Ind.). The active ingredient in Intrepid® 2F pesticide is methoxyfenozide. The methoxyfenozide compound was found to enhance RF2a expression by interaction with the expressed VGE receptor and chimeric promoter comprising Gal4 DNA binding sites described above. Accordingly, the methoxyfenozide compound served as an expression-inducing ligand, which functioned to enhance expression of the RF2a-encoding sequence. Upon application of the methoxyfenozide compound, RF2a expression was induced, thereby allowing RF2a to interact with the Box II-containing E fragment, which was operably linked to the uidA sequence.

More particularly, the plant leaf tissues mentioned above were submerged in stain solution containing 1 mM of X-Glue in 100 mM sodium phosphate buffer (pH 7.0), 2 mM K4Fe(CN)6, 2 mM K3Fe(CN)6, 0.1% Triton X-100 and 20% methanol (Petruccelli et al. 2001). In order to evaluate GUS activity, several substrates are available. The most commonly used are 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) and 4-methyl-umbelliferyl-glucuronide (MUG). The reaction of GUS with X-Glue generates a blue color that is useful in histochemical detection of uidA gene activity. uidA activity (GUS expression) is shown in the “Treated” leaf in FIG. 21.
by a grey, shaded color (compared to the predominately white, "Untreated" leaf). For quantification purposes, MUG is preferred, because the umbelliferyl radical emits fluorescence under UV stimulation, thus providing better sensitivity and easy measurement by fluorometry.

Following vacuum infiltration of the previously submerged leaf tissues, the tissues were incubated at 37°C overnight before washing with 70% (v/v) ethanol. In approximately 70% of the sixty-seven (67) plants analyzed, a constitutive GUS expression pattern was observed in the methoxyfenozide-treated leaves. See FIG. 21 for a subjective comparison of GUS expression in "Treated" and "Untreated" leaf tissue.

Quantitative analysis of GUS activity in methoxyfenozide-treated plants was then conducted. T2 generation transgenic Arabidopsis plants from twelve T1 primary lines were selected on hygromycin selection medium. The lines are individually referred to herein as EGaV-3, EGaV-5, EGaV-17, EGaV-31, EGaV-50, EGaV-51, EGaV-56, EGaV-59, EGaV-63, EGaV-70, EGaV-72, and E:GUS. For each primary line, fourteen 14-day-old T2 plants were used. Nine plants, in three groups, were cultured in MS hydroponic solution containing a 1:8000 dilution of Intrepid 2F®. The remaining nine plants, in three groups, were left untreated and served as controls. After a three-day treatment period, the GUS activity in leaf tissue samples from each plant was quantitated.

Each sample of leaf tissue was ground to a powder in liquid nitrogen. Total protein was extracted from each sample by adding 300 µl of extraction buffer. Protein concentration of each sample was quantitated using a De Protein Assay Kit (Bio-Rad Laboratories, Hercules, Calif.). GUS activity of each sample was quantitated using MUG as substrate and a fluorescence spectrometer (Molecular Devices, Sunnyvale, Calif.) (Petruccelli et al., 2001). Among the twelve lines, the results showed that methoxyfenozide (Intrepid 2F®) treatment induced GUS expression by an average of 2.1 fold in relation to untreated, control plants (FIG. 22A).

Western Blot analysis—To detect the expression of RF2a in select transgenic lines described above, either treated or untreated with Intrepid® 2F, 40 µg of protein of each sample was separated by 12% SDS-PAGE and blotted onto nitrocellulose membrane. The Arabidopsis lines analyzed by Western Blot included EGaV50, E GaV59, E GaV63, E GaV70, E GaV72, and E:GUS. Rabbit anti-RF2a antibodies were first applied to the nitrocellulose membrane, incubated, and washed. Next, anti-rabbit horseradish peroxidase-conjugated secondary antibodies were applied to the membrane, incubated, and washed. Finally, SuperSignal substrate (Pierce, Rockford, Ill.) was applied to the membrane, incubated, and washed.

After the substrate was applied to the membrane, the RF2a bands were revealed (See FIG. 22B, top panel). The presence of RF2a in the leaf tissue samples that were treated with methoxyfenozide (Intrepid 2F®) correlates with the induction of GUS expression observed in similarly treated leaf tissues (described above).

Example 14

Fragments of the Acidic Domain of RF2a

To further demonstrate the versatility of the present gene expression system and, more particularly, the scope of different peptides encompassed by the novel transcription factors of the invention, fragments of the acidic domain of RF2a were used to construct several novel transcription factors. More specifically, transcription factors comprising residues 49 through 116; 49 through 96; 68 through 116; or 68 through 96 of SEQ ID NO:4 (the full-length RF2a transcription factor) were constructed (35S-A-2C7 (comprising the full acidic domain of RF2a)); 35S-A1-2C7, 35S-A2-2C7, and 35S-A3-2C7, respectively).

The foregoing transcription factors further comprised the synthetic 2C7 protein (SEQ ID NO:75)—the synthetic zinc finger DNA-binding domain described above that specifically binds to the 2C7 DNA-binding site (SEQ ID NO:74). The various fragments of the acidic domain of RF2a were fused to the N-terminus of the 2C7 DNA-binding domain. The gene expression cassettes encoding such transcription factors included a CaMV 35S promoter operatively linked to a sequence encoding the respective fragment of the acidic domain, which was operatively linked to a sequence encoding the 2C7 DNA-binding domain (SEQ ID NO:76). The cassettes are referred to in this example collectively or individually as "effector construct."

The various "effector constructs" were prepared using PCR-based procedures and a plasmid (template DNA) containing a coding sequence for the entire RF2a acidic domain (SEQ ID NO:9) operably linked to a coding sequence (SEQ ID NO:76) for the 2C7 zinc finger DNA-binding domain (SEQ ID NO:75). The various fragments of the RF2a acidic domain were amplified from such plasmid using the primer pairs listed in Table 7B (below) which are defined in Table 7A.

TABLE 7A

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>SEQ ID NO</th>
<th>Location within RF2a/Restriction Enzyme Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-2C7 5'</td>
<td>SEQ ID NO:94</td>
<td>Residue 49/BgIII site</td>
</tr>
<tr>
<td>A-2C7 3'</td>
<td>SEQ ID NO:95</td>
<td>Residue 116/BgIII site</td>
</tr>
<tr>
<td>A68-2C7 5'</td>
<td>SEQ ID NO:96</td>
<td>Residue 68/BgII</td>
</tr>
<tr>
<td>A96-2C7 3'</td>
<td>SEQ ID NO:97</td>
<td>Residue 96/BcIII</td>
</tr>
</tbody>
</table>

TABLE 7B

<table>
<thead>
<tr>
<th>Effector Construct</th>
<th>Location within RF2a</th>
<th>5' Primer</th>
<th>3' Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>35S-A-2C7</td>
<td>Residues 49-116</td>
<td>A-2C7 5'</td>
<td>A-2C7 3'</td>
</tr>
<tr>
<td>35S-A2-2C7</td>
<td>Residues 68-116</td>
<td>A-2C7 5'</td>
<td>A-2C7 3'</td>
</tr>
<tr>
<td>35S-A3-2C7</td>
<td>Residues 68-96</td>
<td>A-2C7 5'</td>
<td>A-2C7 3'</td>
</tr>
</tbody>
</table>

The restriction sites listed in Table 7A above, BgIII and BcIII, were used to substitute the various fragments of the acidic domain in place of the full acidic domain. The "reporter construct," C7er2:GUS, carried the uidA coding sequence located downstream of a promoter comprising 6x2C7-binding sites ligated with the minimal promoter of erbB-2 ("er2") (a.k.a. the retinoblastoma minimal promoter). 35S:2C7 encoded the 2C7 protein without an activation domain, i.e., the acidic domain of RF2a or a fragment thereof, and served as a control.

Protoplasts isolated from suspension cultures of BY-2 cells (Nicotiana tabacum L., cv. BrightYellow-2) were transfected via electroporation using procedures well-known in the art. More specifically, the protoplasts were co-transfected with a mixture of DNAs, including 5 µg of reporter construct, 20 µg of a single effector construct (i.e., 35S-A2-2C7, 35S-A1-2C7, 35S-A2-2C7, or 35S-A3-2C7), 5 µg of an internal control plasmid comprising a CaMV 35S promoter operatively.
linked to a GFP-encoding sequence, and 10 µg of herring sperm DNA. The electroporation was conducted using a discharge of 500 µF and 250 V through disposable 0.4 cm cuvettes. Each transient assay was repeated three times per experiment and each experiment was conducted two times.

In addition, electroporation was carried out as described above without effector construct (negative control), without reporter construct (negative control); or, in place of effector construct, 35S-VP16 (a sequence encoding the activation domain of the herpes simplex virus) (positive control).

Quantification of GUS activity in the transfected protoplasts was carried out 24 hours after transfection. More specifically, transfected protoplasts were lysed by freezing and thawing in GUS extraction buffer (pH 7.7), centrifuged, and the supernatants used for GUS enzyme assays. GUS activity was determined by the method of Jefferson et al. (1987). GFP expression (35S-A-2C7) and, more specifically, fragments of the acidic domain comprising residues 49-96 of SEQ ID NO:4 (35S-A-2C7) were measured. This Example further demonstrates that the acidic domain of RF2a, and certain fragments thereof, represent strong activators of gene expression and, furthermore, that the fragment comprising residues 56-84 (relative to the full RF2a protein) produce an activation effect as strong as the full acidic domain.

**Example 15**

Additional Fragments of the Acidic Domain of RF2a

The results of the previous Example 14 were confirmed in this Example 15 and, furthermore, the activity of two additional fragments of the RF2a acidic domain were measured. The 35S-A-2C7, 35S-A1-2C7, 35S-A2-2C7, and 35S-A3-2C7 effector constructs were prepared as described in Example 14 above. The two additional fragments tested in this Example comprised residues 49-84 (referred to herein as "35S-49/84-2C7") and 56-84 (referred to herein as "35S-56/84-2C7") of the full-length RF2a protein (SEQ ID NO:4). The 35S-56/84-2C7 construct was amplified from such plasmid template described in Example 14 using primer pairs represented by SEQ ID NO:94 and 95. The results show that the acidic domain of RF2a is a strong activator of gene expression (35S-A-2C7) and, more specifically, fragments of the acidic domain comprising residues 49-96 of SEQ ID NO:4 (35S-A1-2C7) encompass the majority of such activity.

### TABLE 8

<table>
<thead>
<tr>
<th>Location within RF2a</th>
<th>FIG. 24 Reference</th>
<th>Example 14 Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues 49-116</td>
<td>35S-49/116-2C7</td>
<td>Effector 35S-A2-2C7</td>
</tr>
</tbody>
</table>

*Not shown in the previous Example 14.*

The effector constructs identified in Table 8 were electroporated into BY-2 protoplasts as described in Example 14 above (along with the "reporter constructs" described therein), with the exception that a discharge of 125 µF and 300V (through disposable 0.4 cm cuvettes) was used. Each transient assay was repeated three times per experiment and each experiment was conducted three times. Quantification of GUS activity was carried out as described in Example 14 above.

As shown in FIG. 24(B), the results confirm that the acidic domain of RF2a, and certain fragments thereof, represent strong activators of gene expression and, furthermore, that the fragment comprising residues 56-84 (relative to the full RF2a protein) produce an activation effect as strong as the full acidic domain.
the selection were grown in soil in a growth chamber. Leaves from six week old plants were collected for GUS expression analysis.

The leaves were homogenized, added to GUS extraction buffer, and subsequently centrifuged. The resulting supernatants were used to determine enzymatic activity by the method of Jefferson et al. (1987). The enzyme activity was determined by quantifying fluorescence using a spectrofluorometer (Spectramax Gemini, Molecular Devices Corp., Sunnyvale, Calif.), with a 365 nm excitation wavelength and 455 nm emission wavelength.

As shown in FIG. 26, the various effector constructs activated and/or enhanced GUS expression in the A. thaliana plants (relative to the C7erG/35S-2C7 controls). In addition, FIG. 26 shows that each plant line/effector construct combination produced variable results—albeit well above the expression levels observed in the C7erG/35S-2C7 controls.

<table>
<thead>
<tr>
<th>Construct (FIG. 27)</th>
<th>Promoter</th>
<th>Novel Transcription Factor Sequence</th>
<th>Termination Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGE CsVMV</td>
<td>SEQ ID NO: 107 (V) + SEQ ID NO: 103 (GE)</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>AGE CsVMV</td>
<td>SEQ ID NO: 6 (A) + SEQ ID NO: 103 (GE)</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>A5 x 2GE CsVMV</td>
<td>(2 x SEQ ID NO: 101 (A5) + SEQ ID NO: 103 (GE)</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>A5 x 3GE CsVMV</td>
<td>(3 x SEQ ID NO: 101 (A5) + SEQ ID NO: 103 (GE)</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>A5 x 4GE CsVMV</td>
<td>(4 x SEQ ID NO: 101 (A5) + SEQ ID NO: 103 (GE)</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>AGE x 2 CsVMV</td>
<td>SEQ ID NO: 103 (GE) + (2 x SEQ ID NO: 101 (A5))</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>AGE x 3 CsVMV</td>
<td>SEQ ID NO: 103 (GE) + (3 x SEQ ID NO: 101 (A5))</td>
<td>No ≥ 3</td>
<td></td>
</tr>
<tr>
<td>AGE x 4 CsVMV</td>
<td>SEQ ID NO: 103 (GE) + (4 x SEQ NO: 101 (A5))</td>
<td>No ≥ 3</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 27(A) further shows a diagram of the reporter construct, namely, 5XG-35S:GUS, which was transformed into BY-2 protoplasts along with the effector constructs. As used herein, “5XG” is represented by SEQ ID NO:108 and refers to a pentameric tandem repeat of the nucleic acid sequence that is recognized by, and binds with, the Gal4 protein/domain (from Bakers yeast) (SEQ ID NO:107). “35S” refers to the CaMV 35S promoter (−45/+48). Of course, the GUS-encoding uidA sequence was operably linked to the CaMV 35S promoter.

Protoplasts were co-transfected, by electroporation as described in Example 14 above, with 5 µg of reporter construct DNA, 20 µg of a single effector construct, 5 µg of a 3SS-GFP plasmid, and 10 µg of herring sperm DNA. Methoxyfenozide (“MOF”) was provided to the transformed protoplasts at a final dilution of 1:25,000 (providing a 2.452 µM final concentration of MOF). Quantification of GUS activity from protoplast extracts 24 hours after transfection was performed.

The results of such quantification are shown in FIG. 27(B), which are expressed as the amount of GUS enzyme activity compared with GFP activity—expressed as relative fluorescent units per second. The results are the average of three independent transfections±standard deviation. The presence of MOF is represented in FIG. 27(B) by “I” (wherein “I” means Inducer), whereas the absence of MOF is represented by “–I”. For example, “AGE–I” represents protoplasts transfected with the AGE effector construct and cultured in the presence of MOF, whereas “AGE+I” represents protoplasts transfected with the AGE effector construct and cultured in the absence of MOF.

As shown in FIG. 27(B), the fusion of the acidic domain of RF2a at the amino-terminal position of GE activates GUS expression in the presence of the MOF inducer. Although the activation provided by the RF2a acidic domain (and frag-
RF2a Acidic Domain/Ecdysone Receptor Fusion Proteins in Plants

In this Example, *Arabidopsis thaliana* Col-0 plants were transformed with *A. tumefaciens* bacterium (strain GV3101) containing effector constructs comprising the CsVMV promoter operably linked to a nucleic acid sequence (SEQ ID NO:104) encoding the GE protein, plus (i) a nucleic acid sequence encoding the VP16 acidic activation domain of the herpes simplex virus, (ii) a sequence (SEQ ID NO:9) encoding the full acidic domain at the N-terminus (e.g., AGE proteins) and (iii) a sequence (SEQ ID NO:102) encoding a dimer of the A5 minimal acidic domain ("A5A5"). The foregoing VP16 acidic domain-, RF2a acidic domain-, and A5A5 domain-encoding sequences were oriented to express such domains at the N-terminal of the GE protein. FIG. 28 refers to these constructs as VGE, AGE, and A5AGE, respectively.

The foregoing effector constructs were prepared using standard DNA cloning and PCR procedures, and plasmids containing nucleic acid sequences encoding the VP16 acidic domain, the full length RF2a acidic domain (SEQ ID NO:6), the A5A5 domain (SEQ ID NO:101), and the GE domain (SEQ ID NO:103). Those of ordinary skill in the art will appreciate, however, that such effector sequences (and derivatives thereof) may also be constructed synthetically using standard procedures. The reporter construct described in Example 17 above, 5XG-35S:GUS, was used in this Example.

Table 10 below provides a summary of the various effector constructs, and the amino acid sequences thereof (from the N-to C-terminus), analyzed in this Example.

<table>
<thead>
<tr>
<th>Construct (FIG. 27)</th>
<th>Promoter</th>
<th>Novel Transcription Factor Sequence</th>
<th>Termination Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGE CsVMV SEQ ID NO:109</td>
<td>Nov</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE CsVMV SEQ ID NO:6 (A) +</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5AGE CsVMV (2 x SEQ ID NO:101 (A5)) +</td>
<td>No</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, plasmids containing the effector and reporter constructs described above were transferred by electroporation into *A. tumefaciens* strain GV3101. Agrobacterium isolates that contained the respective plasmids were used to transform *A. thaliana* Col-0 using a standard floral dip method (Clough and Bent, 1998). Seeds from the T1 generation were collected and germinated in Murashige and Skoog medium, containing hygromycin B (30 µg/ml). The seedlings that survived the selection were grown in soil in a growth chamber and allowed to self-fertilize. Next, T2 generation seeds were collected and germinated in Murashige and Skoog medium, which contained hygromycin B (30 µg/ml). The seedlings that survived the selection in a 3:1 segregation pattern were grown in soil in a growth chamber (under 12 hours of light/12 hours of dark per day, at 22° C., 50% humidity).

Analysis of T2 Generation Plants—In this Example, 22 T2 plants were analyzed for methoxyfenozide (MOF) induced gene expression. More particularly, leaf samples from each of the 22 plants were taken before and after induction with MOF (MOF was applied by drenching the soil with a 1:10,000 dilution of MOF (61.3 µM of MOF) for 60 hours). The leaves were homogenized and added to GUS extraction buffer. The supernatants were analyzed for GUS enzymatic activity as described in Example 16 above.

FIG. 28 shows the relative GUS activity of each plant grown in the presence of MOF compared to the relative GUS activity of each plant grown in the absence of MOF. As shown in FIG. 28, the fusion of the acidic domain of RF2a to the amino-terminal position of the GE protein ("AGE"; middle graph), or the presence of two tandem repeats of A5 ("ASASGE"; bottom graph), activated GUS expression in the presence of the MOF inducer as strong as VP16 ("VGE"; top graph).

Analysis of T2 Generation Plants—A total of 16 different plants from a T2 generation plant line (having a segregation of the reporter/effector constructs of 3:1), wherein each plant contained a VGE, AGE, or ASASGE effector construct, were induced with MOF and analyzed for GUS activity (as described above). Referring to FIG. 29, the VGE construct was tested using plant line #11, the AGE construct was tested using plant line #11, and the ASASGE construct was tested using plant line #18. As shown in FIG. 29, the fusion of the acidic domain of RF2a to the amino-terminal position of the GE protein ("AGE"; middle graph), or the presence of two tandem repeats of A5 ("ASASGE"; bottom graph), activated GUS expression (in the T2 generation plants) in the presence of the MOF inducer as strong as VP16 ("VGE"; top graph).

The many aspects and benefits of the invention are apparent from the detailed description, and thus, it is intended for the following claims to cover all such aspects and benefits of the invention which fall within the scope and spirit of the invention. In addition, because numerous modifications and variations will readily occur to those skilled in the art, the claims should not be construed to limit the invention to the exact construction and operation illustrated and described herein. Accordingly, all suitable modifications and equivalents should be understood to fall within the scope of the invention as claimed herein.

REFERENCES


<400> SEQUENCE: 1
ccagtg gcc 15
cctgg

<210> SEQ ID NO 2
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 2
ccagtg gcc 15
cctgg

<210> SEQ ID NO 3
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 3
ggagt gcc 15
ccttc

<210> SEQ ID NO 4
<211> LENGTH: 368
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 4
Met Asn Arg Glu Lys Ser Pro Ile Pro Gly Asp Gly Gly Asp Gly Leu
Pro Pro Gin Ala Thr Arg Arg Ala Gly Pro Pro Ala Ala Ala Ala
Ala Glu Tyr Asp Ile Ser Arg Met Pro Asp Phe Pro Thr Arg Arg Pro
Gly His Arg Arg Ala His Ser Glu Ile Leu Ser Leu Pro Glu Asp Leu
Asp Leu Cys Ala Gly Gly Gly Gly Gly Pro Ser Leu Ser Asp Glu
Asn Asp Glu Leu Phe Ser Met Phe Leu Asp Val Glu Lys Leu Asn
Ser Thr Cys Gly Ala Ser Ser Glu Ala Glu Ala Glu Ser Ser Ser Ala
Ala Ala His Gly Ala Arg Pro Lys His Gin His Ser Leu Ser Met Asp
Glu Ser Met Ser Ile Lys Ala Glu Leu Val Gly Ala Ser Pro Gly
Thr Glu Gly Met Ser Ser Ala Glu Ala Lys Lys Ala Val Ser Val
Lys Leu Ala Glu Leu Ala Leu Asp Pro Lys Arg Ala Lys Arg Ile
Trp Ala Asp Arg Gin Ser Ala Arg Ser Lys Gly Arg Lys Met Arg
Tyr Ile Ala Glu Leu Gly Arg Lys Val Gin Thr Leu Gin Thr Glu Ala
Thr Thr Leu Ser Ala Gln Leu Ala Leu Leu Gln Arg Asp Thr Ser Gly
210 215 220
Leu Thr Thr Glu Asn Ser Glu Leu Arg Leu Leu Leu Gln Thr Met Glu
225 230 235 240
Gln Gln Val His Leu Gln Asp Ala Leu Asn Asp Thr Leu Lys Ser Glu
245 250 255
Val Gln Arg Leu Lys Val Ala Thr Gly Gln Met Ala Asn Gly Gly Gly
260 265 270
Met Met Met Met Phe Gly Gly Met Met Phe Gly Gly Asn Gln
275 280 285
Gln Met Phe Gln Asn Asn Gln Ala Met Gln Ser Met Leu Ala Ala His
290 295 300
Gln Leu Gln Gln Leu Gln Leu His Pro Gln Ala Gln Gln Gln Val
305 310 315 320
Leu His Pro Gln His Gln Gln Gln Pro Leu His Pro Leu Gln Ala
325 330 335
Gln Gln Leu Gln Ala Ala Arg Asp Leu Lys Met Lys Ser Pro Met
340 345 350
Gly Gly Gln Ser Gln Trp Gly Asp Gly Gly Ser Gly Ser Ser Gly Asn
355 360 365

<210> SEQ ID NO 5
<211> LENGTH: 329
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 5

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

<210> SEQ ID NO 6
<211> LENGTH: 67
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
<400> SEQUENCE: 6
Met Gly His Arg Arg Ala His Ser Glu Ile Leu Ser Leu Pro Glu Asp 1 5 10 15
Leu Asp Leu Cys Ala Ala Gly Gly Gly Asp Gly Pro Ser Leu Ser Asp 20 25 30
Glu Asp Leu Gly Leu Phe Met Phe Leu Asp Val Glu Lys Leu 35 40 45
Asn Ser Thr Cys Gly Ala Ser Gin Gin Asp Pro Gin Ser Ser Ser 50 55 60
Ala Ala Ala 65

<210> SEQ ID NO 7
<211> LENGTH: 1107
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
<400> SEQUENCE: 7
atgaacaggg agaaatcccc gatcccggga gacggcggcg acgggttgcc gccgcaggcc 60
acccgccggg cgggccctcc ggcggcggcg gcggcggcgg agtacgacat cagccgcatg 120
cggatttcc cgacgaggaa ccccggccac aggcgcgccc actccgagat cctgagcctc 180
ccccgaggacc tcgacctgtg cgcggccggc ggcggcgacg ggccgtcgct gtcggacgag 240
aacgacgagg agctcttctc catgttcctc gacgtggaga agctgaacag cacgtgcggg 300
gcgtcgtcgg aggcggaggc ggagtcgtcg tccgccgccg cccatggagc gaggccgaag 360
caccagcaca gcctgtccat ggatgagtcg atgtcgatca aggctgagga gctcgtcggg 420
gcgtcgcccg ggacggaggg gatgtcgtcg gcgcgtcggc agaaggcgcgt gtcgcgcggtc 480
aagctcgccg agcttgctct cgtcgatccc aagagggcga aaaggatttg ggctaacaga	 540
caatctgcgg caagatcaaa ggaaaggaaa atgctgaact ttcgagctca gctagcagag	 600
gtcaaaccgc gcagttgctgc gctgatagc tgcagctggc gcgtcagctt cgtcagcctc	 660
gataccagtc gcagttgctgc gctgatagc tgcagctggc gcgtcagctt cgtcagcctc	 720
cgaactgcgg actgcttcgg ccgcttcggc actgcttcgg ccgcttcggc actgcttcggc	 780
aagctgcaag cccgtcagat gcccgggacc ggcggccgag cggccgggag cggccgggag	 840
caccactgg catgctgggc gcggccgggg gcggccgggg gcggccgggg gcggccgggg	 900
cgggctgggaa gccgtagctc gcagttgctgc gctgatagc tgcagctggc gcgtcagctt	 960
cgcgatgtg ccggatagc gcagttgctgc gctgatagc tgcagctggc gcgtcagctt cgtcagcctc	 1020
cgccgactgc ggcggccgg gcggccgggg gcggccgggg gcggccgggg gcggccgggg	 1080
<210> SEQ ID NO 8
<211> LENGTH: 990
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
<400> SEQUENCE: 8
atgcaggagc caaagcacac cgatccggcg gccatgcgcg gggcgcacca ccgccgggcc 60
agatcgaggc cgatccggcg gccatgcgcg gggcgcacca ccgccgggcc 120
ggctgagtc ccctcggcg ccctcggcg ccctcggcg ccctcggcg ccctcggcg ccctcggcg 180
atctcatcc cggccggccc cggccggccc cggccggccc cggccggccc cggccggccc cggccggccc 240
tcgccagcc ccgagcagc ccgagcagc ccgagcagc ccgagcagc ccgagcagc ccgagcagc 300
cgcgcggag ccgcgaggcc ccgcgaggcc ccgcgaggcc ccgcgaggcc ccgcgaggcc ccgcgaggcc 360
accacagc ggtctgctgc gcagttgctgc gcagttgctgc gcagttgctgc gcagttgctgc gcagttgctgc 420
ggcagagcc aactgctgcgg gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 480
ggctgctgctgc gactgctgcgg gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 540
ttctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc
-continued

```
atgggccaca ggcgcgccca ctccgagatc ctgagcctcc ccgaggacct cgacctgtgc
 gcggccggcg gcggcgacgg gccgtcgctg tcggacgaga acgacgagga gctcttctcc
120
atgttcctcg acgtggagaa gctgaacagc acgtgcgggg cgtcgtcgga ggcggaggcg
gagtcgtcgt ccgccgccgc c
201
<210> SEQ ID NO 10
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 10
ccttttatat actggagttc attatacact cttac
35
<210> SEQ ID NO 11
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 11
gtaagagtgt ataatgaact ccagtatata aggag
35
<210> SEQ ID NO 12
<211> LENGTH: 209
<212> TYPE: DNA
<213> ORGANISM: Rice tungro bacilliform virus
<400> SEQUENCE: 12
acgaatcaat aaagaaggac cagaagatat aaagctggaa catcttcaca tgctaccaca
tggctagcat ctttacttta gcatctctat tattgtaaga gtgtataatg accagtgtgc
ccctggactc cagtatataa ggagcaccag agtagtgtaa tagatcatcg atcaagcaag
cgagagcta aacctttaag agagcaaga
209
<210> SEQ ID NO 13
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 13
gatttcacgg gttggggttt cta
23
<210> SEQ ID NO 14
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 14
tcattccagg gcacacactgg atacactctt acaataa
37
<210> SEQ ID NO 15
<211> LENGTH: 37
```
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 15

tgtatccagt gtgcccctgg aatgaactcc agtatat

<210> SEQ ID NO 16
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 16

atacaccagg ggcacactgg ctcttacaat aatagag

<210> SEQ ID NO 17
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 17

aagagccagt gtgcccctgg tgtataatga actccag

<210> SEQ ID NO 18
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 18

caataccagg ggcacactgg atagagatgc taaagta

<210> SEQ ID NO 19
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 19

tctatccagt gtgcccctgg tattgtaaga gtgtata

<210> SEQ ID NO 20
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 20

atagaccagg ggcacactgg gatgctaaag taaagat
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 21

gcatccagt gtcgccctgg tctattattg taagagt 37

SEQ ID NO 22
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 22
tgtacccag ggcacactgg gcacatgtg tgcagt 37

SEQ ID NO 23
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 23
atggccagt gtcgccctgg tagcatcttt actttag 37

SEQ ID NO 24
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 24
gcatccagg ggcacactgg tgaagatgtt ccagctt 37

SEQ ID NO 25
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 25
cttcaccag ggcacactgg tgaagatggt ccagct 37

SEQ ID NO 26
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 26
tgtacccag ggcacactgg cttotattg gattctg 37

SEQ ID NO 27
LENGTH: 37
TYPE: DNA
ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 27
agaagccagtgtgccctgg gaccagaaga tataaag

<210> SEQ ID NO 28
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 28
tgatcaaagcttccagtgtg cccctggacg aatcaataaa gaag

<210> SEQ ID NO 29
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 29
tgatcaaagcttccagtgtg cccctggacg aatcaataaa gaag

<210> SEQ ID NO 30
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 30
cctgtgccagg gcacactgg gtgctcctta tatactg

<210> SEQ ID NO 31
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 31
g cacccagtgtgccctgg gcagtagtg taataga

<210> SEQ ID NO 32
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 32
ttacaccagg gcacactgg ctactcttggt gctcctt

<210> SEQ ID NO 33
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 33

agtagccagt gtgccctgg tgaatatagc atcgat

<210> SEQ ID NO 34
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 34

ttgcccaag ggccacggcag tgaatgatg tctatta

<210> SEQ ID NO 35
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 35

gatcaccagt gtgccctgg agcaagcgag agctcaa

<210> SEQ ID NO 36
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 36

gatcaccagt gtgccctgg agcaagcgag agctaa

<210> SEQ ID NO 37
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 37

tgatacaagc ttccagtgtg cccctggatc cttcgcaaga ccct

<210> SEQ ID NO 38
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 38

tgatacaagc ttccagtgtg cccctggatc cttcgcaaga ccct
created by the inventors.

<400> SEQUENCE: 39
tgatcaagc ttccagtctg cccctgggac gcacaatccc acta 44

<210> SEQ ID NO 40
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 40
tgatcaagc ttccagtctg cccctggggg atgacgcaca atcc 44

<210> SEQ ID NO 41
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 41
tgatcaagc ttccagtctg cccctgggat tccactgacg taag 44

<210> SEQ ID NO 42
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 42
tgatcaagc ttccagtctg cccctggtg atatctccac tgac 29

<210> SEQ ID NO 43
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 43
tgatcaagc ttatccttcg caagaccct 29

<210> SEQ ID NO 44
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 44
tgatcaagc ttgggatgac gcacaatcc 29

<210> SEQ ID NO 45
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
<400> SEQUENCE: 45
tgatcaagg tgtgtgatat tccactgac

<210> SEQ ID NO 46
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 46
gccgcccata tgtgaagat gaacagggag aaatcc

<210> SEQ ID NO 47
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 47
gccgcccata tgtgagggag cgc

<210> SEQ ID NO 48
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 48
gccgcccata tgtggcagg cgc

<210> SEQ ID NO 49
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 49
cgcggatcct cagttgccgc tgcttcctga

<210> SEQ ID NO 50
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 50
cgcggatcct cagtgtggca tgccaccgaa

<210> SEQ ID NO 51
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
SEQ ID NO: 51
LENGTH: 23
TYPE: DNA
ORGANISM: Artificial
FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQ ID NO: 52
LENGTH: 44
TYPE: DNA
ORGANISM: Artificial
FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQ ID NO: 53
LENGTH: 48
TYPE: PRT
ORGANISM: Artificial
FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQ ID NO: 54
LENGTH: 87
TYPE: PRT
ORGANISM: Artificial
FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQ ID NO: 55
LENGTH: 97
TYPE: PRT
ORGANISM: Artificial
FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.
<210> SEQ ID NO 56
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 56
gccagatcta tgggccacag gcgcgccc

<210> SEQ ID NO 57
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 57
cgcggatcct ccatgggcgg cggcg

<210> SEQ ID NO 58
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 58
gccagatcta tgaacaggga gaaatcccc

<210> SEQ ID NO 59
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 59
cgcggatccg gggttcctcg tcgggaa

<210> SEQ ID NO 60
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 60
gccagatcta tgaattcgg aggcaaccag c

<210> SEQ ID NO 61
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 61
cgcggatccg ttgcccctgc ttctgta
<210> SEQ_ID NO 62
<211> LENGTH: 775
<212> TYPE: DNA
<213> ORGANISM: Rice tungro bacilliform virus

<400> SEQUENCE: 62
agatcttctta caaaagggag tagtaatatt taatgagctt gaaggaggat atcaactctc 60
tccaggtttt attgagaccc ttattgtcct ttgttttat aaaaaaatt aaccttacaac 120
caggtctct gaaggctac cgccaatctg tgccagaatga taatgttcc 180
tgatcctaact ttcttctttata tagaaggaga ggctcttcgtg ggcacctcctc tgaataagaa 240
aagagaaca cagatgagga ccacatattata tacttctcaag aattatctgct gtctctatgg 300
tattcaaca ttaagggcat acaacatctaat aaccttggaa gtatttgggt ttgatttacta 360
catttgggaa atatgggttg atgaaagaaga aggagaggttg atgatactt ccagattatac 420
taatgattaca aagagatcta aagagagcagtg agggctttgg gegaacagatg atgagaacac 480
aatgtcctcg accacatcata tagaagatgtaa atgtatcatctt caaccgagctcc 540
agcttttact aaccagcaat aaccattctg acaattaaag aaagaggaaac agatattaaa 600
cggaacatc ttaacagtgt accacatggcg gctcatactttttta attttctatgtttaatt 660
gtgaaggtgt ataatgcaccacct ttctggcccct ggaacctcagta atataaggag cccagatga 720
gtgaataaga tctgatcata agcaagagag aagtaaactt aatagagagag caaga 775

<210> SEQ_ID NO 63
<211> LENGTH: 209
<212> TYPE: DNA
<213> ORGANISM: Rice tungro bacilliform virus

<400> SEQUENCE: 63
acgaatcaat aaagaaggac cagaagatat aaagctggaa catcttcaca tgctaccaca 60
tggctagcat ctttacttta gcatctctat tattgtaaga gtggtataatg accagtgtgg 120
cgtgagctc cagatattata ggaagatcag aagattgttaa tagaagcagta ataaagcaag 180
cgagagctca aacctttcaag agagaaga 209

<210> SEQ_ID NO 64
<211> LENGTH: 209
<212> TYPE: DNA
<213> ORGANISM: Rice tungro bacilliform virus

<400> SEQUENCE: 64
acgaatcaat aaagaaggac cagaagatat aaagctggaa catcttcaca tgctaccaca 60
tgctagcatttttctctat tattgtaagta gtattgtaatg accagtgtgg 120
cgatcagctc cagatattata ggaagatcag aagattgttaa tagaagcagta ataaagcaag 180
cgagagctca aacctttcaag agagaaga 209

<210> SEQ_ID NO 65
<211> LENGTH: 194
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE: OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 65
acgaatcaat aaagaaggac cagaagatat aaagctggaa catcttcaca tgctaccaca 60
tgctagcatttttctctat tattgtaagta gtattgtaatg accagtgtgg 120
tataaggagc accagagtag tgaatagat cactgatcaag caagcagaga gcctaaactt
180
tagagagc aaga
194

<210> SEQ ID NO 66
<211> LENGTH: 108
<212> TYPE: DNA
<213> ORGANISM: Cauliflower mosaic virus

<400> SEQUENCE: 66
ttgatgtgat atctccactg acgtaaggga tgacgcacaa tcccactatc cttcgcaaga
60
cctcctct atataaggaa gttcatttca tttggagagg acacgctg
108

<210> SEQ ID NO 67
<211> LENGTH: 552
<212> TYPE: DNA
<213> ORGANISM: Cassava vein mosaic virus

<400> SEQUENCE: 67
ccaagatgta gcatcaagaa tccaatgttt acgggaaaaa ctatggaagt attatgtgag
60
cgcagcaaga agcagatcaa tatgcggcac atatgcaacc tattttcaaa aatgaagaat
120
getacagatac aagatcctat actgccagaa tacgaagaag aatacgtaga aattgaaaaa
180
gagagaaccctg caagcagaga ggaagtacat atatgcaacc tattttcaaa aatgaagaat
240
getacagatac aagatcctat actgccagaa tacgaagaag aatacgtaga aattgaaaaa
300
gagagaaccctg caagcagaga ggaagtacat atatgcaacc tattttcaaa aatgaagaat
360
gagagaaccctg caagcagaga ggaagtacat atatgcaacc tattttcaaa aatgaagaat
420
gagagaaccctg caagcagaga ggaagtacat atatgcaacc tattttcaaa aatgaagaat
480
tctcgagaa gttctcctct ctctctctct ttggtgaggg acacgctg
540

<210> SEQ ID NO 68
<211> LENGTH: 320
<212> TYPE: PRT
<213> ORGANISM: Artificial

<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
 created by the inventors.

<400> SEQUENCE: 68
Met Gly His Arg Arg Ala His Ser Glu Ile Leu Ser Leu Pro Glu Asp
1  5  10  15
Leu Asp Leu Cys Ala Ala Gly Gly Gly Gly Gly Gly Pro Ser Leu Ser Asp
20 25 30
Glu Asn Asp Glu Leu Phe Ser Met Phe Leu Asp Val Glu Lys Leu
35 40 45
Asn Ser Thr Cys Gly Ala Ser Ser Glu Ala Glu Ala Glu Ser Ser Ser
50 55 60
Ala Ala Ala His Gly Ala Arg Pro Gly His Glu His Ser Leu Ser Met
65 70 75 80
Asp Glu Ser Met Ser Ile Ala Glu Leu Val Gly Ala Ser Pro Gly
85 90 95
Thr Glu Gly Met Ser Ser Ala Glu Ala Lys Lys Ala Val Ser Ala Val
100 105 110
Lys Leu Ala Glu Leu Ala Leu Val Asp Pro Lys Arg Ala Lys Arg Ile
115 120 125
Trp Ala Asn Arg Glu Ser Ala Ala Arg Ser Lys Glu Arg Lys Met Arg
130 135 140
Tyr Ile Ala Glu Leu Glu Arg Lys Val Gin Thr Leu Gin Thr Glu Ala
145 150 155 160
Thr Thr Leu Ser Ala Gln Leu Ala Leu Leu Gin Arg Asp Thr Ser Gly
165 170 175
Leu Thr Thr Glu Asn Ser Glu Leu Lys Leu Arg Leu Gin Thr Met Glu
180 185 190
Gln Gln Val His Leu Gln Asp Ala Leu Asp Thr Leu Lys Ser Glu
195 200 205
Val Gin Arg Leu Lys Val Ala Thr Gly Gin Met Ala Asn Gly Gly Gly
210 215 220
Met Met Met Asn Phe Gly Gly Met Pro His Gin Phe Gly Gly Ann Gin
225 230 235 240
Gln Met Phe Gin Asn Gin Gln Ala Met Gin Ser Met Leu Ala Ala His
245 250 255
Gln Leu Gin Gln Leu Gin Leu His Pro Gin Ala Gin Glu Gin Glu Val
260 265 270
Leu His Pro Gin His Gin Gln Gin Gin Gin Gin Gin Gin Gin Gin Gin
275 280 285
Gln Gin Leu Gin Gln Ala Arg Asp Leu Lys Met Lys Ser Pro Met
290 295 300
Gly Gly Gin Ser Gin Trp Gly Asp Gly Lys Ser Gly Ser Gly Asn
305 310 315 320

<210> SEQ ID NO: 69
<211> LENGTH: 283
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 69
Met Asn Arg Glu Lys Ser Pro Ile Pro Gly Asp Gly Asp Gly Leu
1 5 10 15
Pro Pro Gin Ala Thr Arg Arg Ala Gly Pro Pro Ala Ala Ala Ala
20 25 30
Ala Glu Tyr Asp Ile Ser Arg Met Pro Asp Phe Pro Thr Arg Asn Pro
35 40 45
Gly His Arg Arg Ala His Ser Glu Ile Leu Ser Leu Pro Glu Asp Leu
50 55 60
Asp Leu Cys Ala Ala Gly Gly Gly Asp Gly Asp Ser Leu Ser Asp Leu
65 70 75 80
Asn Asp Glu Glu Leu Phe Ser Met Phe Leu Asp Val Glu Lys Leu Asn
85 90 95
Ser Thr Cys Gly Ala Ser Ser Glu Ala Glu Ala Glu Ser Ser Ser Ala
100 105 110
Ala Ala His Gly Ala Arg Pro Lys His Gin His Ser Leu Ser Met Asp
115 120 125
Glu Ser Met Ser Ile Lys Ala Glu Leu Val Gly Ala Ser Pro Gly
130 135 140
Thr Glu Gly Met Ser Ser Ala Glu Ala Lys Ala Val Ser Ala Val
145 150 155 160
Lys Leu Ala Glu Leu Ala Val Asp Pro Lys Arg Ala Lys Arg Ile
Trp Ala Aen Arg Gln Ser Ala Ala Arg Ser Lys Glu Arg Lys Met Arg

Tyr Ile Ala Glu Leu Glu Arg Lye Val Gln Thr Leu Gin Thr Glu Ala

Thr Thr Leu Ser Ala Gin Leu Ala Leu Leu Gin Arg Asp Thr Ser Gly

Leu Thr Thr Glu Aen Ser Glu Leu Lys Leu Arg Leu Gin Thr Met Glu

Gln Gin Val His Leu Gin Asp Ala Leu Aen Asp Thr Leu Lys Ser Gin

Val Gin Arg Leu Lys Val Ala Thr Gly Gin Met Ala Aen Gly Gin Gly

Met Met Met Aen Phe Gly Gly Met Pro His Gin
Met Lys Ser Pro Met Gly Gly Gln Ser Gln Trp Gly Asp Gly Lys Ser
245 250 255
Gly Ser Ser Gly Asn
260

<210> SEQ ID NO 71
<211> LENGTH: 236
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 71

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

<210> SEQ ID NO 72
<211> LENGTH: 176
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 72

Met Ser Ser Ser Ala Ala His Gly Ala Arg Pro Lys His Gin His
1  5   10   15
Ser Leu Ser Met Asp Glu Ser Met Ser Ile Lys Ala Glu Glu Leu Val
20  25  30
**US 7,847,064 B2**

**-continued**

<table>
<thead>
<tr>
<th>Gly Ala Ser Pro Gly Thr Glu Gly Met Ser Ser Ala Glu Ala Lys Lys</th>
<th>35</th>
<th>40</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala Val Ser Ala Val Lys Leu Ala Glu Leu Ala Leu Val Asp Pro Lys</td>
<td>50</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Arg Ala Lys Arg Ile Trp Ala Asn Arg Gln Ser Ala Ala Arg Ser Lys</td>
<td>65</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>Glu Arg Lys Met Arg Tyr Ile Ala Glu Leu Glu Arg Lys Val Gln Thr</td>
<td>85</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Leu Gln Thr Glu Ala Thr Thr Leu Ser Ala Gln Leu Ala Leu Leu Gln</td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Arg Asp Thr Ser Gly Leu Thr Thr Glu Asn Ser Gln Leu Lys Leu Arg</td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Leu Gln Thr Met Glu Glu Gln Val His Leu Gln Asp Ala Leu Asn Asp</td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Thr Leu Lys Ser Glu Val Gln Arg Leu Lys Val Ala Thr Gln Met</td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Ala Asn Gly Gly Gly Met Met Met Asn Phe Gly Gly Met Pro His Gln</td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>&lt;210&gt; SEQ ID NO 73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;211&gt; LENGTH: 23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;212&gt; TYPE: DNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;213&gt; ORGANISM: Artificial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;220&gt; FEATURE:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;223&gt; OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| ttcacacagg aaacagctat gac                                       | 23 |

|<210> SEQ ID NO 74                                              |    |    |    |
|<211> LENGTH: 18                                               |    |    |    |
|<212> TYPE: DNA                                                |    |    |    |
|<213> ORGANISM: Artificial                                    |    |    |    |
|<220> FEATURE:                                                |    |    |    |
|<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors. |

| gcgtgggccg gcgtgggccg | 18 |

|<210> SEQ ID NO 75                                              |    |    |    |
|<211> LENGTH: 172                                               |    |    |    |
|<212> TYPE: PRT                                                |    |    |    |
|<213> ORGANISM: Artificial                                    |    |    |    |
|<220> FEATURE:                                                |    |    |    |
|<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors. |

<table>
<thead>
<tr>
<th>Met Leu Glu Leu Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phe Ser Lys Ser Ala Asp Leu Lys Arg His Ile Arg Ile His Thr Gly</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asp Phe Ser Arg Ser</td>
<td>35</td>
<td>40</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Asp His Leu Thr Thr His Ile Arg Thr His Thr Gln Lys Pro Phe</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg Lys</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
</tr>
</tbody>
</table>
Arg His Thr lys Ile His Thr Gly Glu Lys Pro Tyr Ala Cys Pro Val
85 90 95
Glu Ser Cys Asp Arg Arg Phe Ser Lys Ser Ala Asp Leu Lys Arg His
100 105 110
Ile Arg Ile His Thr Gly Gin Lys Pro Phe Gin Cys Arg Ile Cys Met
115 120 125
Arg Asn Phe Ser Arg Ser Asp His Ile Thr Thr His Ile Arg Thr His
130 135 140
Thr Gly Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala
145 150 155 160
Arg Ser Asp Glu Arg Lys Arg His Thr Lys Ile His
165 170

<210> SEQ ID NO 76
<211> LENGTH: 516
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 76
atgctcgagc tcccctatgc ttgccctgtc gagtcctgcg atcgccgctt ttctaagtcg
60
gctgatctga agcgccatat ccgcatccac acaggccaga agcctttcca gtgtcgaata
120
tgcatgcgta acttcagtcg tagtgaccac cttaccaccc acatccgcac ccacacagcc
180
gagaagcctt ttgcctgtga catttgtggg aggaagtttg ccaggagtga tgaacgcaag
240
aggcatacca aaatccatac cggtgagaag ccctatgctt gccctgtcga gtcctgcgat
300
cgccgctttt ctaagtcggc tgatctgaag cgccatatcc gcatccacac aggccagaag
360
cccttccagt gtcgaatatg catgcgtaac ttcagtcgta gtgaccacct taccacccac
420
atccgcaccc acacaggcga gaagcctttt gcctgtgaca tttgtgggag gaagtttgcc
480
aggagtgatg aacgcaagag gcataccaaa atccat
516

<210> SEQ ID NO 77
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 77
ccccaaagtc cagcttgaaa t
21

<210> SEQ ID NO 78
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.
<400> SEQUENCE: 78
ttaatccaac ttggaaaatg
20

<210> SEQ ID NO 79
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 79
caccacccccc

SEQ ID NO: 80
LENGTH: 15
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 80
caccaccaccctac

SEQ ID NO: 81
LENGTH: 12
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 81
tggaccctacca

SEQ ID NO: 82
LENGTH: 13
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 82	aggtcaccccatt

SEQ ID NO: 83
LENGTH: 18
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 83
tggatggtggaagacagca

SEQ ID NO: 84
LENGTH: 65
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

SEQUENCE: 84

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

<210> SEQ ID NO 85
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 85
caacgcgtca cgtggcctac a
21

<210> SEQ ID NO 86
<211> LENGTH: 62
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 86
Lys Ala Glu Arg Lys Arg Met Arg Asn Arg Ile Ala Ala Ser Lys Ser
1 5 10 15
Arg Lys Arg Lys Leu Glu Arg Ile Ala Arg Leu Glu Glu Lys Val Lys
20 25 30
Thr Leu Lys Ala Gln Asn Ser Glu Leu Ala Ser Thr Ala Asn Met Leu
35 40 45
Arg Gln Val Ala Gln Leu Lys Gln Lys Val Met Asn His
50 55 60

<210> SEQ ID NO 87
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 87
cgtcgatgac gtcatcgacg
20

<210> SEQ ID NO 88
<211> LENGTH: 90
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or
created by the inventors.

<400> SEQUENCE: 88
Met Glu Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe
1 5 10 15
Ser Gln Lys Thr Asn Leu Asp Thr His Ile Arg Ile His Thr Gly Gln
20 25 30
Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Glu His Thr
35 40 45
Gly Leu Asn Gln His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala
50 55 60
Cys Asp Ile Cys Gly Arg Lys Phe Ala Thr Leu His Thr Arg Asp Arg
His Thr Lys Ile His Leu Arg Gln Lys Asp
85 90

<210> SEQ ID NO 89
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 89

gacgctataa aaggag

<210> SEQ ID NO 90
<211> LENGTH: 76
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 90

Pro Pro Gly Thr Pro Ser Arg His Glu Lys Ser Leu Gly Leu Leu Thr
1 5 10 15
Thr Lys Phe Val Ser Leu Leu Gln Glu Ala Lys Asp Gly Val Leu Asp
20 25 30
Leu Lys Leu Ala Ala Asp Thr Leu Ala Val Arg Gln Lys Arg Arg Ile
35 40 45
Tyr Asp Ile Thr Asn Val Leu Glu Gly Ile Gly Leu Ile Gly Lys Lys
50 55 60
Ser Lys Asn Ser Ile Gln Trp Lys Gly Val Gly Pro
65 70 75

<210> SEQ ID NO 91
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 91

attttcgctgc ggtttt

<210> SEQ ID NO 92
<211> LENGTH: 163
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 92

gctcgagaga cagtactcgc ctcgaggac agtactcgc ctcgaggaca gatactcgc
60
cgagacagt ctcgagtcgc gcggagcagct ctcgagtcgc tcaactgcgc gacaacctcc
120
tctatataag gaagttcatt tcatttggag aggacacgct gaa
163

<210> SEQ ID NO 93
<211> LENGTH: 2595
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: This sequence was artificially derived and/or created by the inventors.

<400> SEQUENCE: 93

atggcrococg cacgcttgag gcggacccatc tccagcagc cggagcagct 60
gcgcctgacg atgcacagc gacgtacagc tccgatgcca cgcgtccgag 120
ttcctccccacgc gcggcgttcg gcgtccgctg cctgagcagc gcggagcagc 180
gacgtggcgt tggacctgc gctgtcgagc ctgaggtcag ccgggttcag 240
cctttccgagat ctcgatgcca cgcgtccgtgc gcgtcagcgc ctgaggtcag 300
tccgctgccg aacgtagctc tctggcagc gcgtgacgct gcgtcagcgc 360
cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 420
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 480
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 540
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 600
gcgtgacgct gcgtcagcgc ccctgacgct gcgtgacgct gcgtcagcgc 660
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 720
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 780
gcgtgacgct gcgtcagcgc ccctgacgct gcgtgacgct gcgtcagcgc 840
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 900
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 960
cagacacac tatccagcgtt ccctgacgct gcgtgacgct gcgtcagcgc 1020
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1080
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1140
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1200
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1260
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1320
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1380
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1440
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1500
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1560
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1620
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1680
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1740
agcgtgtgctt ccagcttgcctg ccctgacgct gcgtgacgct gcgtcagcgc 1800
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1860
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1920
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 1980
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 2040
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 2100
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 2160
ttcgctgctgag cctgacggtcg cgcattgcg cgcgtgctgg cgtgacgct gcgtcagcgc 2220
gttcctgcctcg cgaacaagcta cgcgacaact acggcaagcc tggcatggcc 2280
taacctgcgcgctgcgtg gctctgcgtct accttttctt a cgcgcaaggg tggagagcag 2340
atccattacg cgctgctcagcgctgtcgtc accttttctt a cgcgcaaggg tggagagcag 2400
gcgcacatggctaacgacgtacatgtgcctgtcgtctgctgtgcatgt tacctatggc 2460
gttaacagctgcagtcgcctgcagtccgtcatgcgtcgtcatgcgtcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtcatgcgtcgtc
<210> SEQ ID NO 99
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Primer used to amplify sequence from Rice (Oryza sp.)

<400> SEQUENCE: 99

cgaggatcccttcggctgtc 27

<210> SEQ ID NO 100
<211> LENGTH: 584
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Chimeric VGE receptor, comprising the acidic
domain of the SV40 virus (Simian virus 40), the Gal4 protein from
Baker’s yeast (Salmonella sp.), and the ecdysone receptor of
Cloristomeura fumiferana

<400> SEQUENCE: 100

Met Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His Leu Asp
 1   5    10   15
Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Phe Asp
 20   25   30
Leu Asp Met Leu Gly Asp Gly Ser Pro Gly Pro Gly Phe Thr Pro
 35   40   45
His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe Glu Phe
 50   55   60
Glu Glu Met Phe Thr Asp Ala Leu Gly Ile Asp Gly Tyr Gly Gly Lys
 65   70   75   80
Leu Leu Gly Thr Ser Arg Arg Ile Ser Gly Gly Phe Gly Gly Met
 85   90   95
Lys Leu Leu Ser Ser Ile Glu Ala Cys Asp Ile Cys Arg Leu Lys
100  105  110
Lys Leu Lys Cys Ser Lys Lys Gly Pro Lys Cys Ala Lys Cys Leu Lys
115  120  125
Asn Asn Trp Glu Arg Tyr Ser Pro Lys Thr Arg Ser Pro Leu
130  135  140
Thr Arg Ala His Leu Thr Glu Val Gly Ser Arg Leu Arg Gly Leu Gly
145  150  155  160
Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile Leu
165  170  175
Lys Met Asp Ser Leu Glu Asp Ile Lys Ala Leu Thr Gly Leu Phe
180  185  190
Val Glu Asp Asn Val Asn Lys Arg Ala Val Thr Asp Arg Leu Ala Ser
195  200  205
Val Glu Thr Asp Met Pro Leu Thr Arg Glu His Arg Ile Ser Ala
210  215  220
Thr Ser Ser Ser Glu Ser Ser Asn Lys Gly Glu Arg Glu Leu Thr
225  230  235  240
Val Ser Gly Ser Arg Ile Ser Thr Pro Glu Cys Val Val Pro
245  250  255
Glu Thr Gin Cys Ala Met Lys Arg Lys Glu Lys Ala Gin Lys Glu
260  265  270
Lys Asp Lys Leu Pro Val Ser Thr Thr Val Asp Asp His Met Pro
Pro Ile Met Gln Cys Glu Pro Pro Pro Pro Glu Ala Ala Arg Ile His
275  280  285
Glu Val Val Pro Arg Phe Leu Ser Asp Lys Leu Leu Glu Thr Asn Arg
290  295  300
305
Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn Gin Gin Phe Leu Ile Ala
310  315  320
325
Arg Leu Ile Trp Tyr Gin Asp Gly Tyr Glu Gin Pro Ser Asp Glu Asp
330  335  340
345
Leu Lys Arg Ile Thr Gin Thr Thr Gin Gin Ala Asp Asp Glu Asn Glu
350  355  360
365
Glu Ser Asp Thr Pro Phe Arg Gin Ile Thr Glu Met Thr Ile Leu Thr
370  375  380
Val Gin Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Gly Phe Ala Lys
385  390  395
400
Ile Ser Gin Pro Asp Gin Ile Thr Leu Leu Lys Ala Phe Ser Ser Glu
405  410  415
Val Met Met Leu Arg Val Ala Arg Arg Tyr Asp Ala Ala Ser Asp Ser
420  425  430
435
440
Val Leu Phe Ala Asn Asn Gin Ala Tyr Thr Arg Asp Asn Tyr Arg Lys
445
Ala Gly Met Ala Tyr Val Ile Glu Leu Leu His Phe Cys Arg Cys
450  455  460
Met Tyr Ser Met Ala Leu Asp Asn Ile His Tyr Ala Leu Leu Thr Ala
465  470  475
480
Val Val Ile Phe Ser Asp Arg Pro Gly Leu Glu Gin Pro Gin Leu Val
485  490  495
Glu Glu Ile Gln Arg Tyr Tyr Leu Arg Thr Leu Arg Ile Tyr Ile Leu
500  505  510
Asn Gin Leu Ser Gly Ser Ala Arg Ser Ser Val Ile Tyr Gly Lys Ile
515  520  525
530
Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu Gly Met Gin Asn Ser Asn
535  540
545
Met Cys Ile Ser Leu Lys Leu Asn Arg Lys Leu Pro Pro Phe Leu
550  555  560
Glu Glu Ile Trp Asp Val Ala Asp Met Ser His Thr Gin Pro Pro Pro
565  570  575
580
Ile Leu Glu Ser Pro Thr Asn Leu

<210> SEQ ID NO 101
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: Oryza sp.

<240> SEQUENCE: 101
Glu Ile Leu Ser Leu Pro Glu Leu Asp Leu Cys Ala Ala Gly Gly
1  5  10  15
Gly Asp Gly Pro Ser Leu Ser Asp Gin Gin Gin Gin
20  25

<210> SEQ ID NO 102
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Oryza sp.
gagatcctga gacctcccg gaacctccag ctgtgcgcgg ccggcggcgg cgacgggccg
tcgtgcgag acgagaacga cggagag

Lys Leu Leu Ser Ser Ile Glu Gin Ala Cys Asp Ile Cys Arg Leu Lys
Lys Leu Lys Cys Ser Lys Glu Glu Pro Lys Cys Ala Lys Cys Leu Lys
Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro Leu
Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu Glu
Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile Leu
Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu Phe
Val Gln Asp Ann Val Ann Lys Asp Ala Val Thr Asp Arg Leu Ala Ser
Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gin His Arg Ile Ser Ala
Thr Ser Ser Ser Glu Ser Ser Ann Lys Gly Gin Arg Gin Leu Thr
Val Ser Gly Gly Ser Arg Ile Ser Trp Pro Glu Cys Val Val Pro
Glu Thr Gin Cys Ala Met Lys Arg Lys Glu Lys Ala Gln Lys Glu
Lys Asp Lys Leu Pro Val Ser Thr Thr Thr Val Asp Asp His Met Pro
Pro Ile Met Gln Cys Glu Pro Pro Pro Pro Glu Ala Ala Arg Ile His
Glu Val Val Pro Arg Phe Leu Ser Arg Lys Leu Leu Glu Thr Asn Arg
Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn Gln Gin Phe Leu Ile Ala
Arg Leu Ile Trp Gin Asp Gly Tyr Glu Gin Pro Ser Asp Glu Asp
Leu Lys Arg Ile Thr Gln Thr Trp Gin Gin Ala Asp Glu Asn Glu
Glu Ser Asp Thr Pro Phe Arg Gin Ile Thr Glu Met Thr Ile Leu Thr
Val Gln Leu Ile Val Gin Ala Lys Gly Leu Pro Gly Phe Ala Lys
Ile Ser Gin Pro Asp Gin Ile Thr Leu Leu Lys Ala Phe Ser Ser Glu
Val Met Met Leu Arg Val Ala Arg Asp Ala Ala Ser Asp Ser
Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr Arg Asp Asn Tyr Arg Lys
340 345 350

Ala Gly Met Ala Tyr Val Ile Glu Leu Leu His Phe Cys Arg Cys
355 360 365

Met Gly Ser Met Ala Leu Arg Asp Ala Tyr Val Thr Ala Leu Leu Thr Ala
370 375 380

Val Val Ile Phe Ser Asp Pro Gly Leu Glu Gin Pro Gin Leu Val
385 390 395 400

Glu Glu Ile Gin Arg Tyr Leu Asn Thr Tyr Ile Tyr Ile Leu
405 410 415

Asn Gly Leu Ser Gly Ser Ala Arg Ser Ser Val Ile Tyr Gly Lys Ile
420 425 430

Leu Ser Ile Leu Ser Glu Arg Thr Leu Gly Met Gin Asn Ser Gin
435 440 445

Met Cys Ile Ser Leu Lys Leu Arg Leu Arg Leu Arg Leu Leu Leu
450 455 460

Glu Glu Ile Trp Asp Val Ala Asp Met Ser His Thr Gln Pro Pro Pro
465 470 475 480

Ile Leu Glu Ser Pro Thr Asp Leu
485

<210> SEQ ID NO 104
<211> LENGTH: 1464
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Nucleic acid sequence encoding SEQ ID NO:103 (a
combination of the Ga14 protein and ecdysone receptor)
<400> SEQUENCE: 104
aagctactgt cttctatcga acaagcatgc gatatttgcc gacttaaaaa gctcaagtgc 60
tccaaagaaa aaccaagtct gcgcatctga actggtgact tcgctactct 120
cccaaaaacc cagagctcgc gctgactagg gcacatctga cagaagtgga atcaaggcta 180
gaaagactgg aacagctatt tctactgatt tttcctcgag aagaccttga catgattttg 240
aaatctggttt cttataacag tataaaacga ttgttaacag gattattttg acaagataat 300
gtaaataag atgcctgacg agatagatgtgtcctcagag atcctcctaca 360
ttgatacggg tataagaagta ctgtgagatg aaagagagag aagagagag 420
agacagttga ctgctactgg ggtcoopagc gcaagctcgg gatcatcggc 480
gagacatctg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg 540
gactgctagg gggctgctgg gggctgctgg gggctgctgg gggctgctgg gggctgctgg 600
gtttaacag atggagctag cagacatcgc ttcctcttcag cagatcgcctag gggctgctgg 660
ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg 720
gtttaacag atggagctag cagacatcgc ttcctcttcag cagatcgcctag gggctgctgg 780
agacagttga ctgctactgg ggtcoopagc gcaagctcgg gatcatcggc 840
gagacatctg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg 900
gactgctagg gggctgctgg gggctgctgg gggctgctgg gggctgctgg gggctgctgg 960
gtttaacag atggagctag cagacatcgc ttcctcttcag cagatcgcctag gggctgctgg 1020
agacagttga ctgctactgg ggtcoopagc gcaagctcgg gatcatcggc 1080
gagacatctg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg ggcctgctgg 1140
ctgctacagc tctgctcat cttttctgac cggccagggt tggagcagcc gcaactggtg
ggaagagatcc agcggtacta cctgaatacg ctccgcatct atatcctgaa ccagctgagc
gggtcggcgc gttcgtccgt catatacggc aagatcctct caatcctctc tgaagctcgc
acgctggcga tcgaaaacctc caacatgtgc atctccctca agctcaagaa cagaaagctg
cgcctttcc tcgaggagat ctgggatgtg gcggacatgt cgcacaccca accgccgct
atcctgagt cccccacgaa tctc

<210> SEQ ID NO 105
<211> LENGTH: 334
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Ecdysone receptor domain from Cloristoneura funiferana

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

<210> SEQ ID NO 106
<211> LENGTH: 1002
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Nucleic acid sequence encoding the ecdysone receptor domain from Clooristoneura fumiferana

<400> SEQUENCE: 106

```
cctgagtgcg tagtacccga gactcagtgc gccatgaagc ggaaagagaa gaaagcacag 60
aagggaag acaaaactgcc tgtcagcagc acaacggttg agacaccaag gcggccatt 120
atcgagtttg aacctcacc ctcctgaagca gcaaggattc acgaagtggt cccaaggttt 180
tcttgcgaca agctgttgga gacaaaccgg cagaaaaaca tcccccagtt gcagagcaac 240
cagcaacccct eogcaggatg cagacacact tctctgcaga gctcatctgg taccaggagc 300
gctcatctgg 360
gagatttga agaggattac gcagacgtgg cagcaagcgg acgatgaaaa cgaagagtct
```

<210> SEQ ID NO 107
<211> LENGTH: 154
<212> TYPE: PRT
<213> ORGANISM: Salmonella sp.

<400> SEQUENCE: 107

```
Lys Leu Leu Ser Ser Ile Glu Gin Ala Cys Asp Ile Cys Arg Leu Lys 1  5  10  15
Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu Lys 20  25  30
Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Arg Ser Pro Leu 35  40  45
Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu Glu 50  55  60
Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile Leu 65  70  75  80
Lys Met Asp Ser Leu Gin Asp Ile Lys Ala Leu Leu Thr G1y Leu Phe 85  90  95
```
What is claimed is:

1. A protein capable of regulating the expression level of at least one gene of interest, which comprises:
   (i) an amino acid sequence consisting of SEQ ID NO:6; and
   (ii) a DNA binding domain selected from the group consisting of a 2C7 DNA binding domain and a GAL4 DNA binding domain, wherein the DNA binding domain is capable of recognizing and binding to a nucleic acid sequence, wherein the nucleic acid sequence is capable of regulating the expression level of at least one gene of interest.

2. The protein of claim 1, wherein the nucleic acid sequence is a promoter.

3. The protein of claim 1, which further comprises an ecdysone receptor domain.

4. The protein of claim 3, wherein the ecdysone receptor domain comprises an amino acid of SEQ ID NO:105.

5. The protein of claim 4, wherein the nucleic acid sequence comprises a Ga4 binding domain, which comprises a nucleic acid sequence of SEQ ID NO:108.

6. The protein of claim 1, wherein the nucleic acid sequence comprises a C7er domain, which comprises a sequence of SEQ ID NO:74.

7. A protein capable of regulating the expression level of at least one gene of interest, which comprises:
   (i) an amino acid sequence that comprises SEQ ID NO:101; and
   (ii) a DNA binding domain selected from the group consisting of a 2C7 DNA binding domain and a GAL4 DNA binding domain, wherein the DNA binding domain is capable of recognizing and binding to a nucleic acid sequence, wherein the nucleic acid sequence is capable of regulating the expression level of at least one gene of interest.

8. The protein of claim 7, wherein the nucleic acid sequence is a promoter.

9. The protein of claim 8, which further comprises an ecdysone receptor domain.

10. The protein of claim 9, wherein the ecdysone receptor domain comprises an amino acid sequence of SEQ ID NO:105.

11. The protein of claim 8, which further comprises a 2C7 zinc finger domain.

12. The protein of claim 11, wherein the promoter comprises a C7er domain, which comprises a nucleic acid sequence of SEQ ID NO:74.
13. The protein of claim 7, which comprises two or more repetitions of SEQ ID NO:101.

14. A protein capable of regulating the expression level of at least one gene of interest, which comprises:
   (i) an amino acid sequence represented by SEQ ID NO:6 or a fragment thereof which consists of SEQ ID NO:101;
   (ii) an ecdysone receptor domain; and
   (iii) a Gal4 protein.

* * * * *