



US 20140335592A1

(19) **United States**

(12) **Patent Application Publication**
Voytas et al.

(10) **Pub. No.: US 2014/0335592 A1**

(43) **Pub. Date: Nov. 13, 2014**

(54) **TAL EFFECTOR-MEDIATED DNA
MODIFICATION**

continuation of application No. 12/965,590, filed on
Dec. 10, 2010, now Pat. No. 8,586,363.

(71) Applicants: **Iowa State University Research
Foundation, Inc.**, Ames, IA (US);
**Regents of the University of
Minnesota**, Minneapolis, MN (US)

(60) Provisional application No. 61/285,324, filed on Dec.
10, 2009, provisional application No. 61/352,108,
filed on Jun. 7, 2010, provisional application No.
61/366,685, filed on Jul. 22, 2010.

(72) Inventors: **Daniel F. Voytas**, Falcon Heights, MN
(US); **Adam J. Bogdanove**, Ithaca, NY
(US); **Feng Zhang**, Plymouth, MN (US)

Publication Classification

(73) Assignees: **Iowa State University Research
Foundation, Inc.**, Ames, IA (US);
**Regents of the University of
Minnesota**, Minneapolis, MN (US)

(51) **Int. Cl.**
C12N 9/22 (2006.01)
(52) **U.S. Cl.**
CPC **C12N 9/22** (2013.01); **C12Y 301/21004**
(2013.01)
USPC **435/196**

(21) Appl. No.: **14/187,713**

(22) Filed: **Feb. 24, 2014**

(57) **ABSTRACT**

Related U.S. Application Data

(63) Continuation of application No. 13/738,381, filed on
Jan. 10, 2013, now Pat. No. 8,697,853, which is a

Materials and Methods related to gene targeting (e.g., gene
targeting with transcription activator-like effector nucleases;
“TALENs”) are provided.

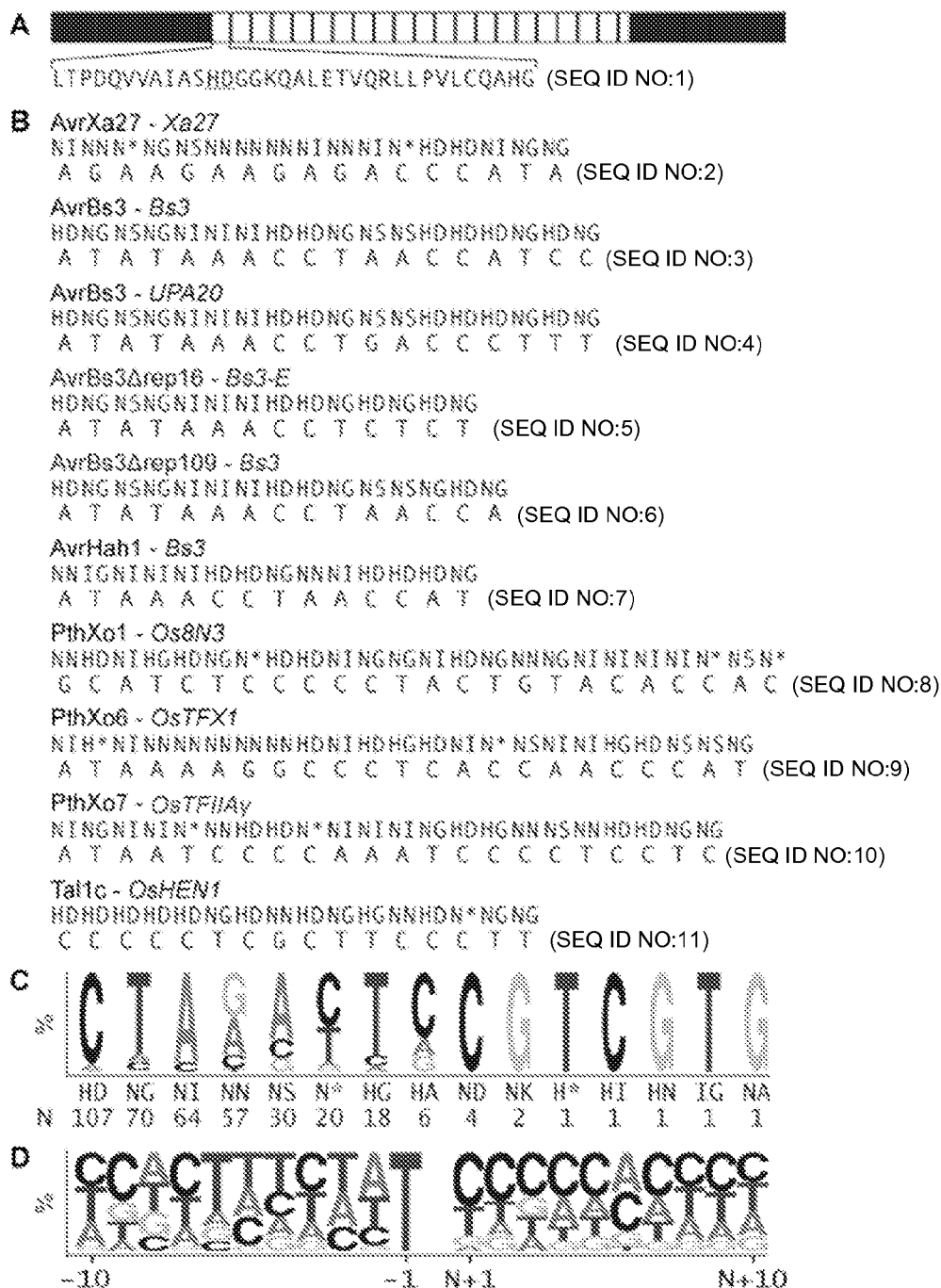
Figure 1

Figure 2

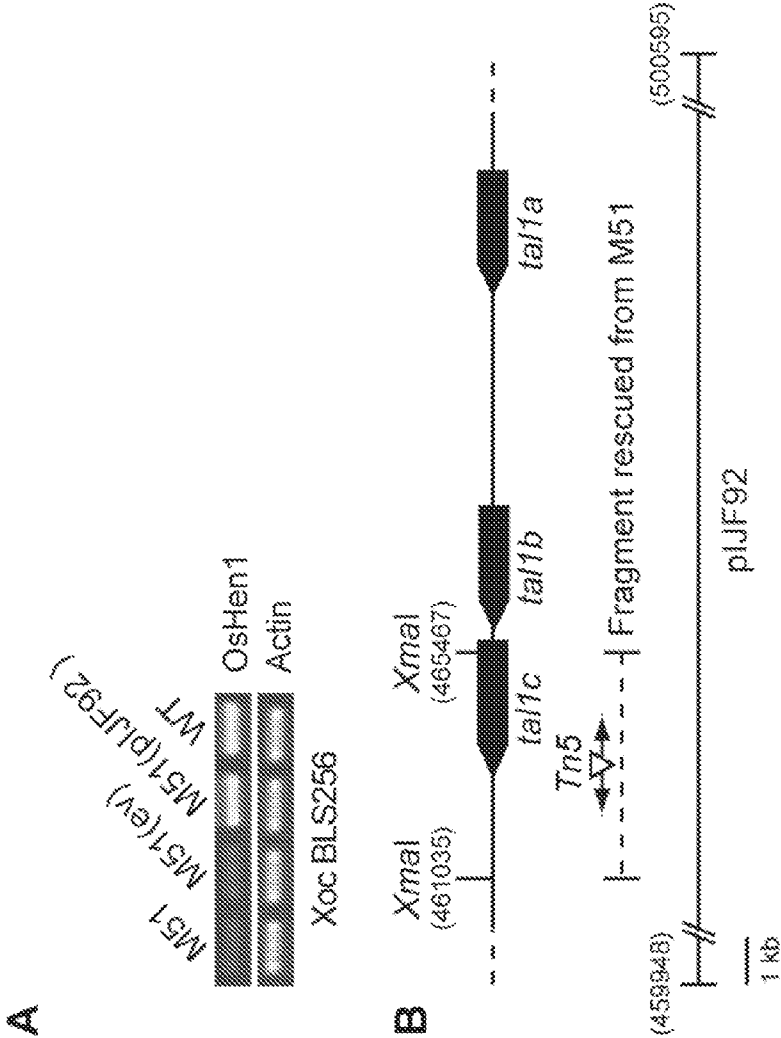


Figure 3

MDPIRSRTPSPARELLPGPQPDGVQPTADRGVSPGAGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQFDPSTLNTSL
FDSLPPFGAHTTEAATGEWDEVQSGLRADAPPPTMRVAVTAARPPRAKPAPRRRAAQPSDASPAQVLDLRTLGYSSQQQKEKI
KPKVRSTVAQHHEALVGHGFTAHIVALSQHPAALGTAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPPL
QLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTPQVVAIASHDGGKQALETVQRLLPVLCCQAHGLTPQQVVAIASNNGGK
QALETVQRLLPVLCCQAHGLTPQQVVAIASNNGGKQALETVQRLLPVLCCQAHGLTPQVVAIASNNGGKQALETVQRLLPVLCCQAH
GLTPQVVAIASNIGGKQALETVQALLPVLCCQAHGLTPQVVAIASNIGGKQALETVQALLPVLCCQAHGLTPQVVAIASNIGGKQA
LETVQALLPVLCCQAHGLTPQVVAIASHDGGKQALETVQRLLPVLCCQAHGLTPQVVAIASHDGGKQALETVQRLLPVLCCQAHGLT
PQQVVAIASNNGGKQALETVQRLLPVLCCQAHGLTPQVVAIASNNGGKQALETVQALLPVLCCQAHGLTPQVVAIASNNGGKQAL
ETVQRLLPVLCCQAHGLTPQVVAIASHDGGKQALETVQRLLPVLCCQAHGLTPQVVAIASHDGGKQALETVQRLLPVLCCQAHGLT
PEQVVAIASHDGGKQALETVQRLLPVLCCQAHGLTPQQVVAIASNNGGGRPALETVQRLLPVLCCQAHGLTPQVVAIASHDGGKQAL
ETVQRLLPVLCCQAHGLTPQQVVAIASNNGGGRPALESIVAQLSRPDPALAAALNDHLVALACLGGRPALDAVKKGLPHAPALIKRTN
RRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTELEARSGTLPPASQRWDRILQASGM
KRAKPSPTSTQTPDQASLHAFADSLERDLDAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQQSFEVRVPEQRDALHPLSWRV
KRPRTSIGGGLPDGTPGTTAADLAASSTVMREQDEDPFAGAADDFPAFNEEELAWLMELLPQ (SEQ ID NO:12)

[illegible]

Figure 5

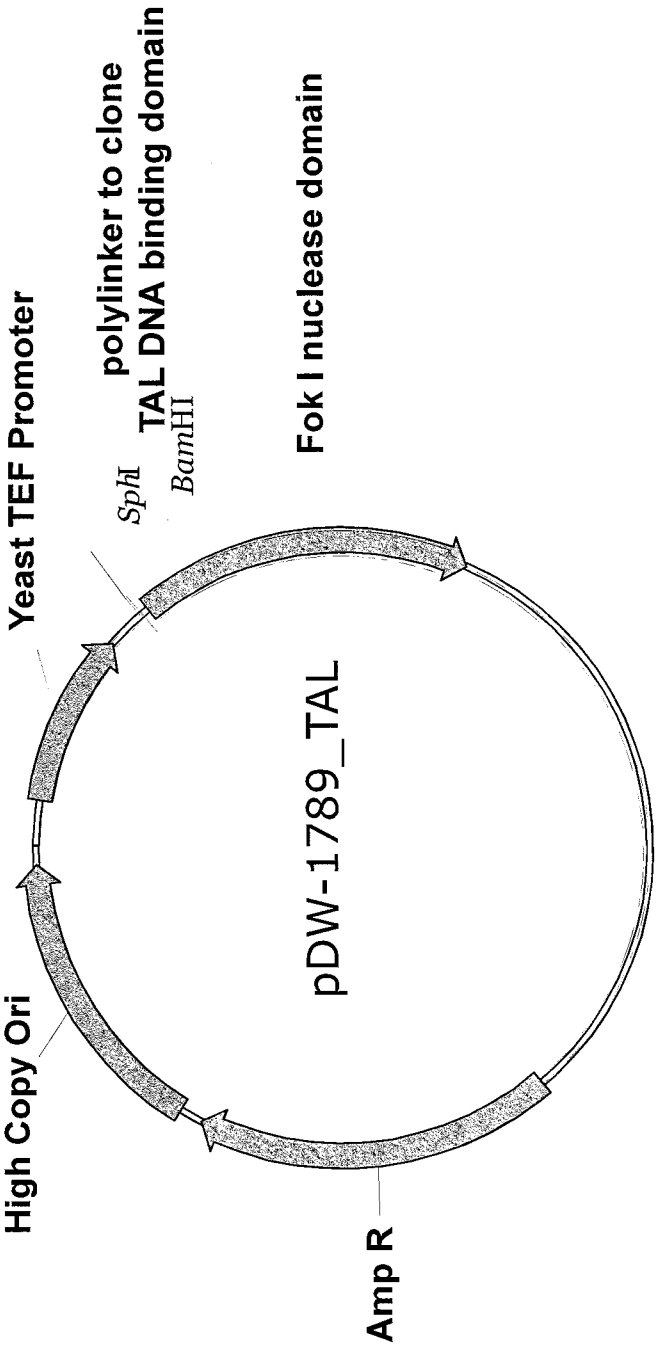


Figure 6

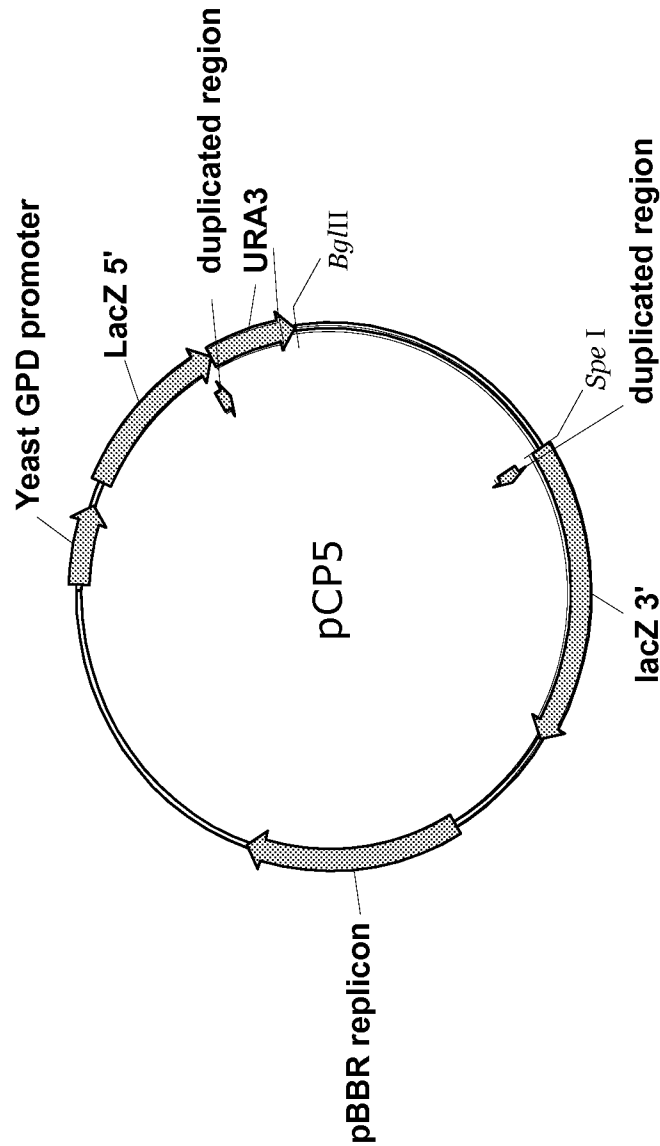


Figure 7

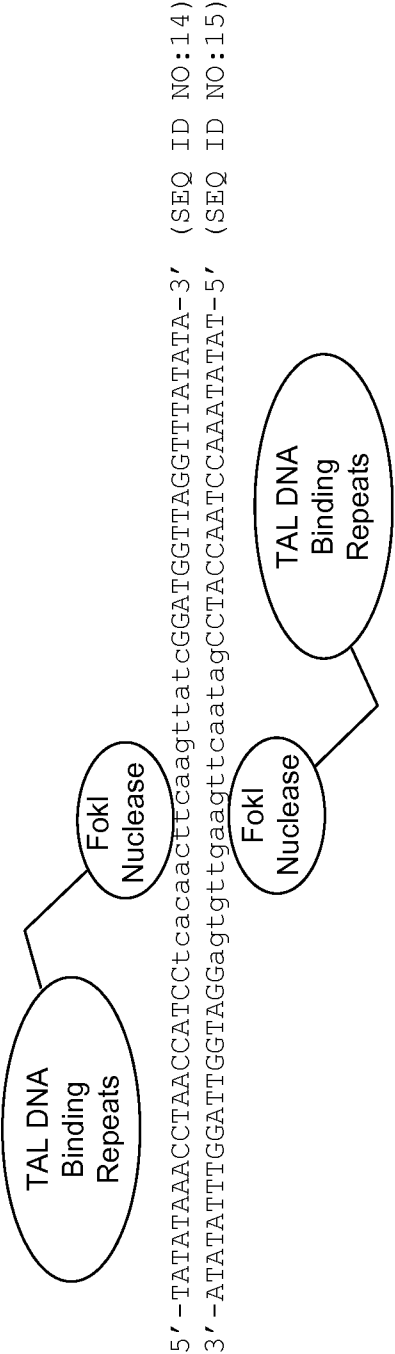


Figure 8

LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPQQVVAIAS	NG	GGKQALETVQRLLPVLCCQAHG
LTPQQVVAIAS	NS	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	NG	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	NI	GGKQALETVQALLPVLCCQAHG
LTPEQVVAIAS	NI	GGKQALETVQALLPVLCCQAHG
LTPEQVVAIAS	NI	GGKQALETVQALLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPQQVVAIAS	NG	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	NS	GGKQALETVQALLPVLCCQAHG
LTPEQVVAIAS	NS	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPQQVVAIAS	NG	GGRPALETVQRLLPVLCCQAHG
LTPEQVVAIAS	HD	GGKQALETVQRLLPVLCCQAHG
LTPQQVVAIAS	NG	GGRPALE (SEQ ID NO:16)

Figure 9

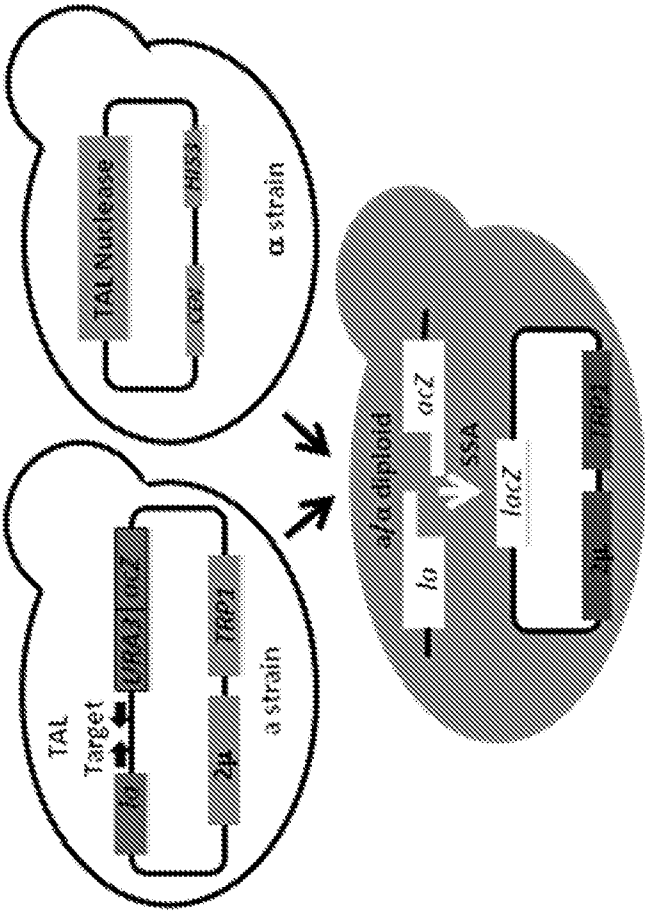


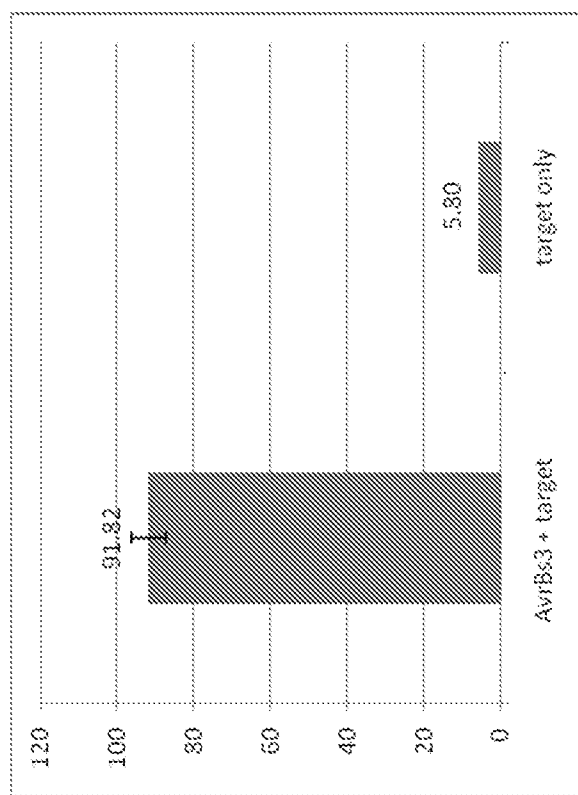
Figure 10

Figure 11

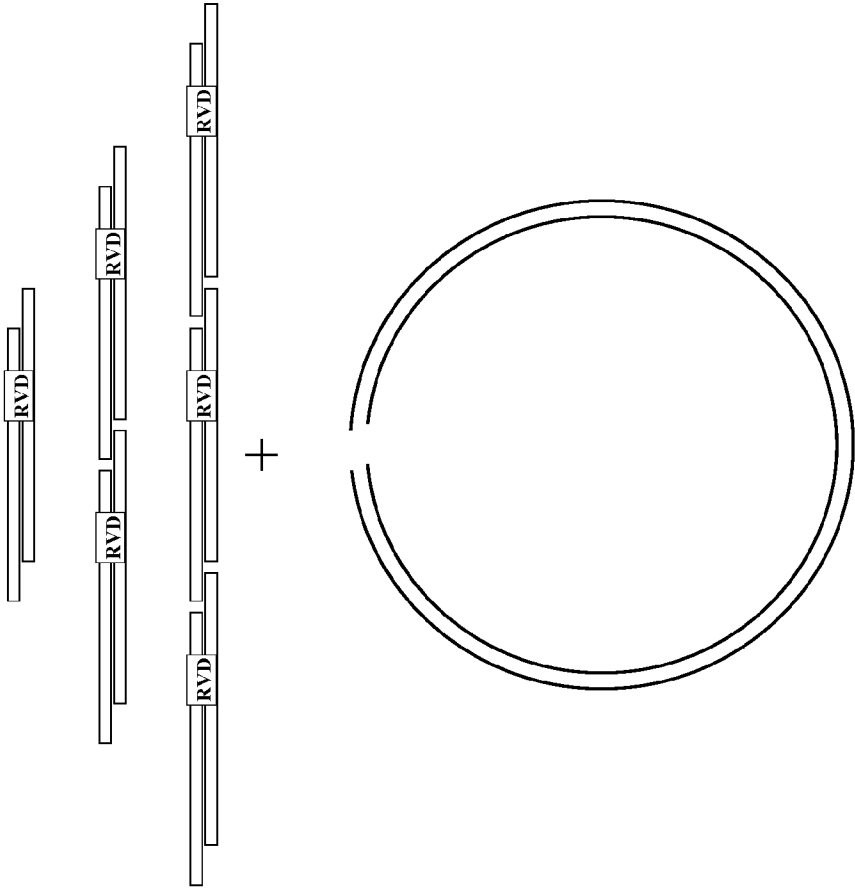


Figure 12

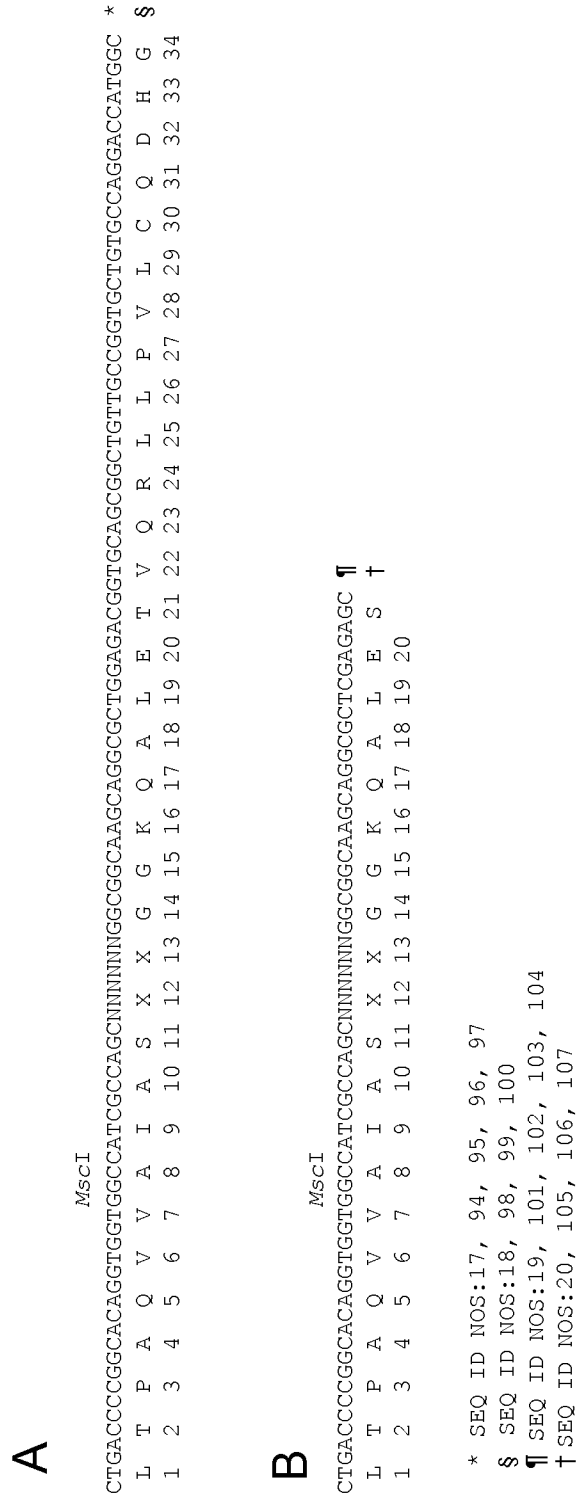


Figure 13

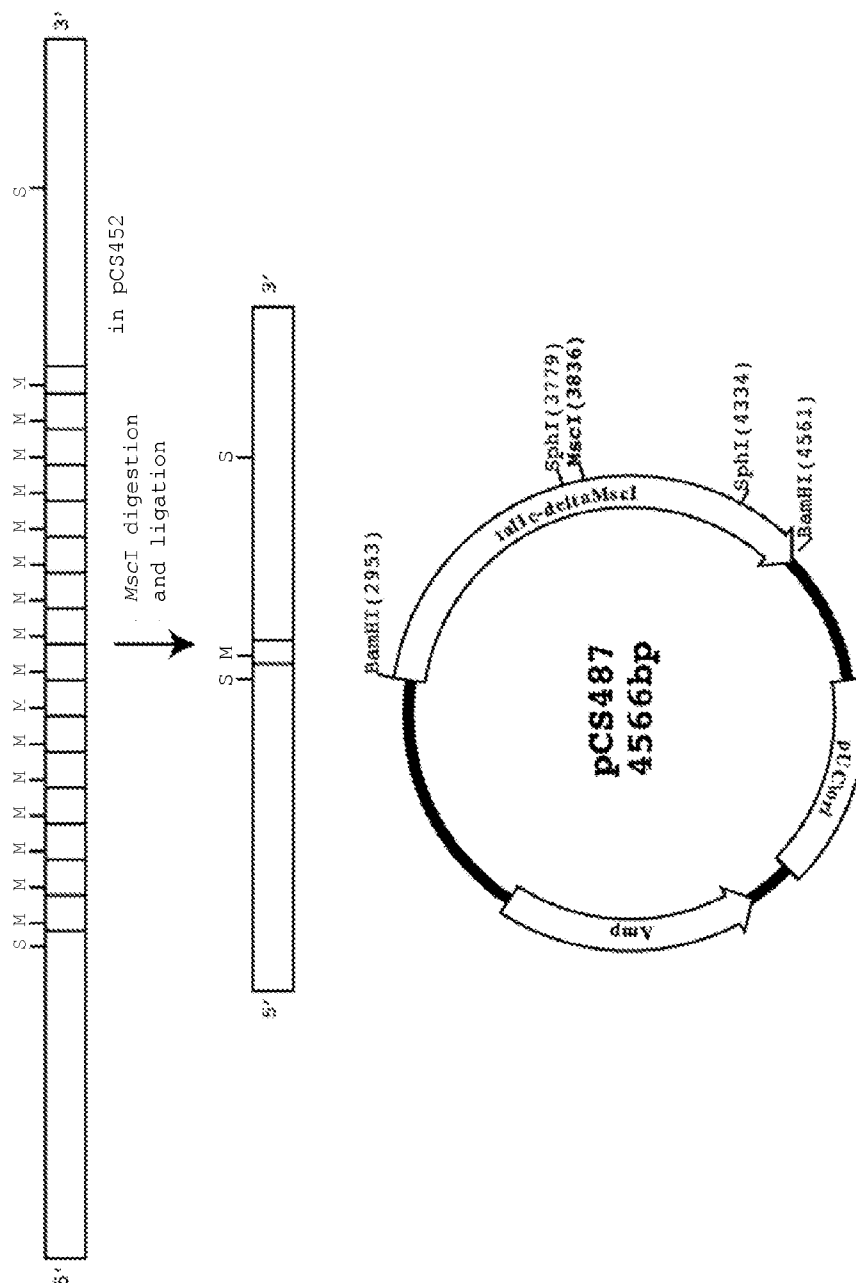


Figure 14

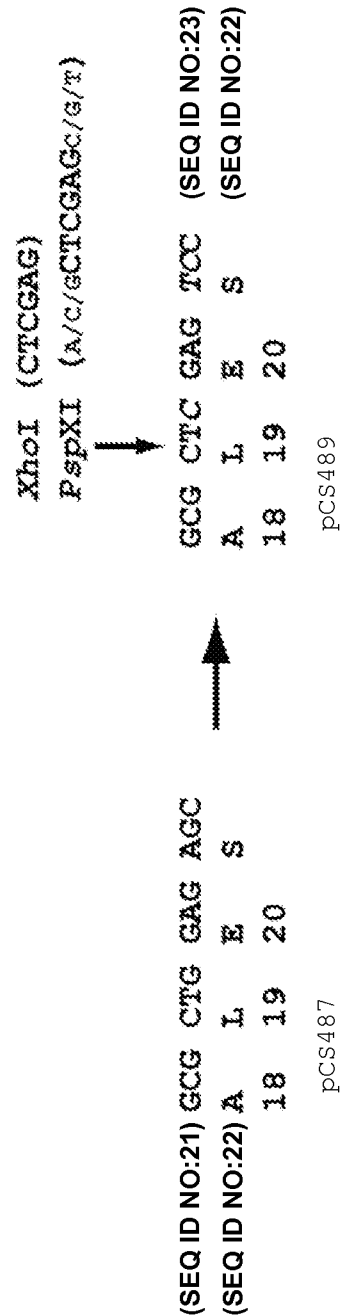


Figure 15

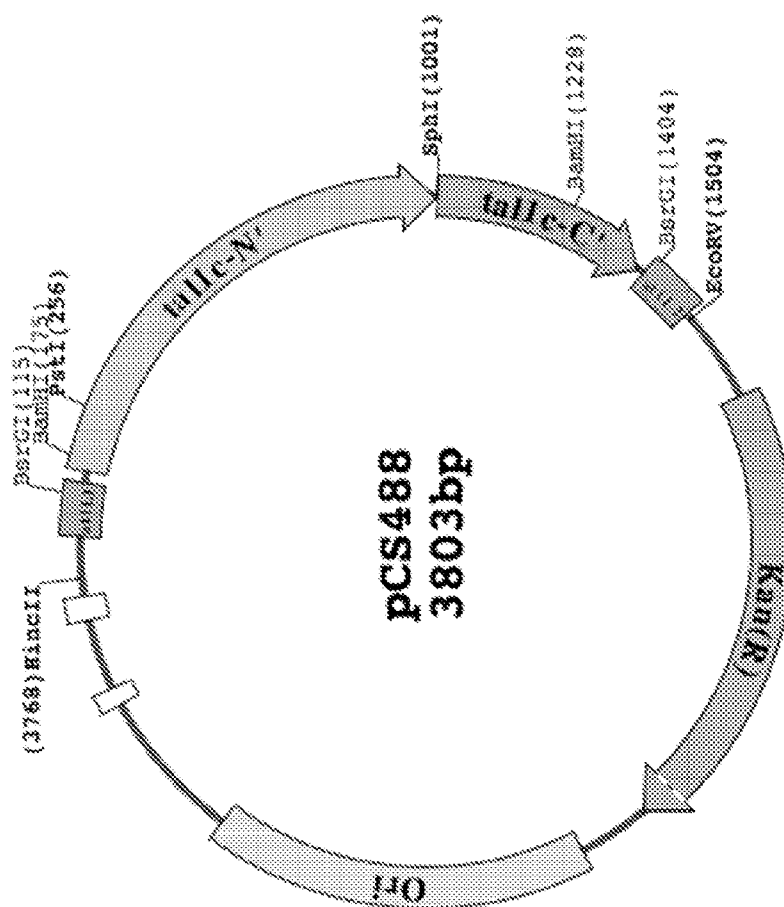


Figure 16

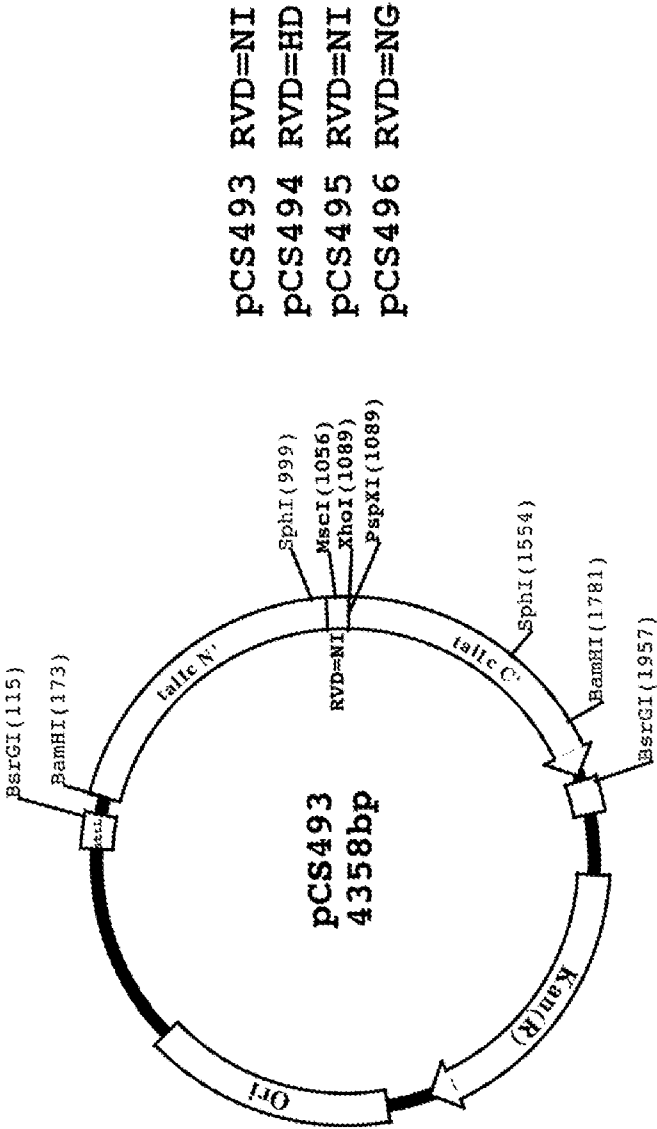


Figure 18

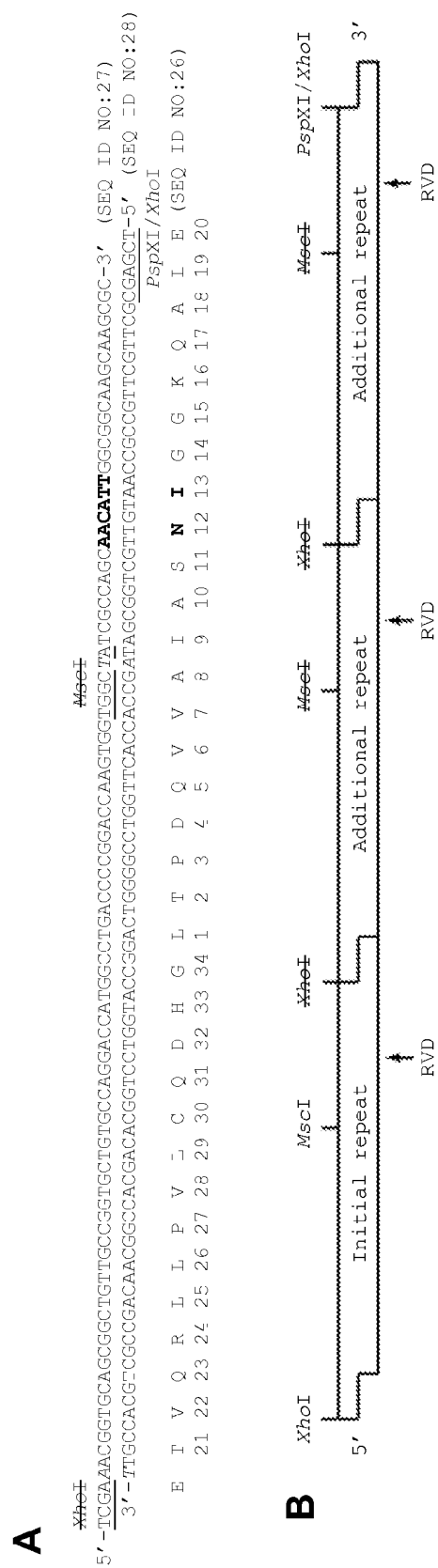


Figure 19

Plasmid	Description
PCS670	NG-NI-NI
PCS671	NG-NI-HD
PCS672	NG-NI-NN
PCS673	NG-NI-NG
PCS674	NG-HD-NI
PCS675	NG-HD-HD
PCS676	NG-HD-NN
PCS677	NG-HD-NG
PCS678	NG-NN-NI
PCS679	NG-NN-HD
PCS680	NG-NN-NN
PCS681	NG-NN-NG
PCS682	NG-NG-NI
PCS683	NG-NG-HD
PCS684	NG-NG-NN
PCS685	NG-NG-NG

Plasmid	Description
PCS686	NN-NI-NI
PCS687	NN-NI-HD
PCS688	NN-NI-NN
PCS689	NN-NI-NG
PCS690	NN-HD-NI
PCS691	NN-HD-HD
PCS692	NN-HD-NN
PCS693	NN-HD-NG
PCS694	NN-NN-NI
PCS695	NN-NN-HD
PCS696	NN-NN-NN
PCS697	NN-NN-NG
PCS698	NN-NG-NI
PCS699	NN-NG-HD
PCS700	NN-NG-NN
PCS701	NN-NG-NG

Plasmid	Description
PCS638	HD-NI-NI
PCS639	HD-NI-HD
PCS640	HD-NI-NN
PCS641	HD-NI-NG
PCS642	HD-HD-NI
PCS643	HD-HD-HD
PCS644	HD-HD-NN
PCS645	HD-HD-NG
PCS646	HD-NN-NI
PCS647	HD-NN-HD
PCS648	HD-NN-NN
PCS649	HD-NN-NG
PCS650	HD-NG-NI
PCS651	HD-NG-HD
PCS652	HD-NG-NN
PCS653	HD-NG-NG

Plasmid	Description
PCS628	NI-NI-NI
PCS629	NI-NI-HD
PCS630	NI-NI-NN
PCS631	NI-NI-NG
PCS632	NI-HD-NI
PCS633	NI-HD-HD
PCS634	NI-HD-NN
PCS635	NI-HD-NG
PCS636	NI-NN-NI
PCS637	NI-NN-HD
PCS638	NI-NN-NN
PCS639	NI-NN-NG
PCS640	NI-NG-NI
PCS641	NI-NG-HD
PCS642	NI-NG-NN
PCS643	NI-NG-NG

Plasmid	Description
PCS602	NI
PCS603	HD
PCS604	NN
PCS605	NG

Plasmid	Description
PCS606	NI-NI
PCS607	NI-HD
PCS608	NI-NN
PCS609	NI-NG
PCS610	HD-NI
PCS611	HD-HD
PCS612	HD-NN
PCS613	HD-NG
PCS614	NN-NI
PCS615	NN-HD
PCS616	NN-NN
PCS617	NN-NG
PCS618	NG-NI
PCS619	NG-HD
PCS620	NG-NN
PCS621	NG-NG

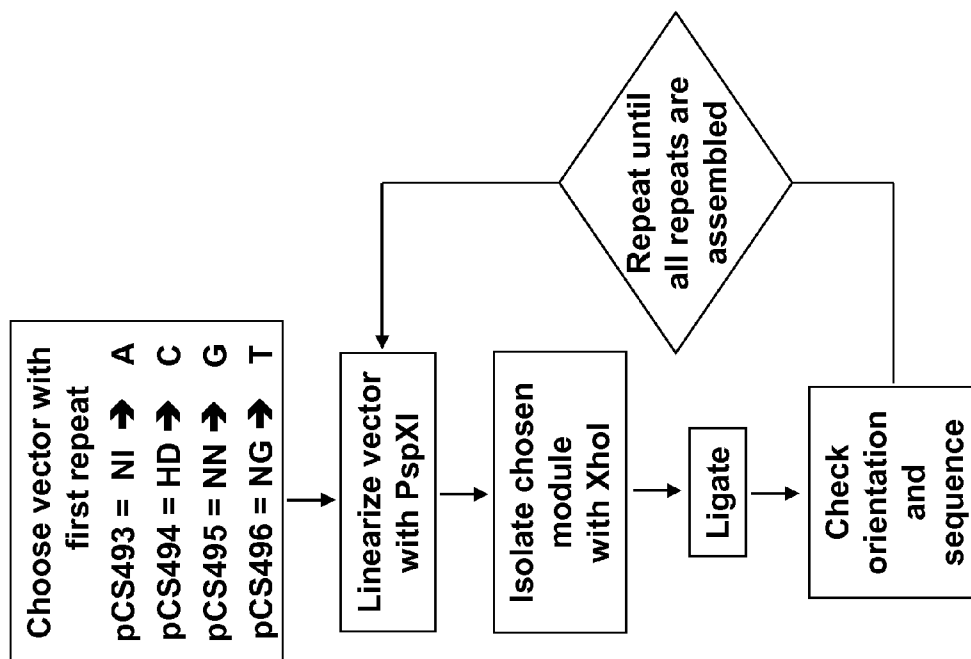
Figure 20

Figure 21B

Adh Target DNA Adh RVDs	C C C A A G A A G T A A A C A T (SEQ ID NO:30) HD HD HD HI NN NI NI NN NG NI NI NI HD NI NG									
pMAT 61 pMAT 62 pMAT 63 pMAT 64	HD HD HD pMAT1		HD HD HD pMAT1		HD HD HD pMAT61		HD HD HD pMAT62		HD HD HD pMAT63	
			NI NN NI 530		NI NN NI 533		NI NI NI 522		HD NI NG 541	

Figure 22

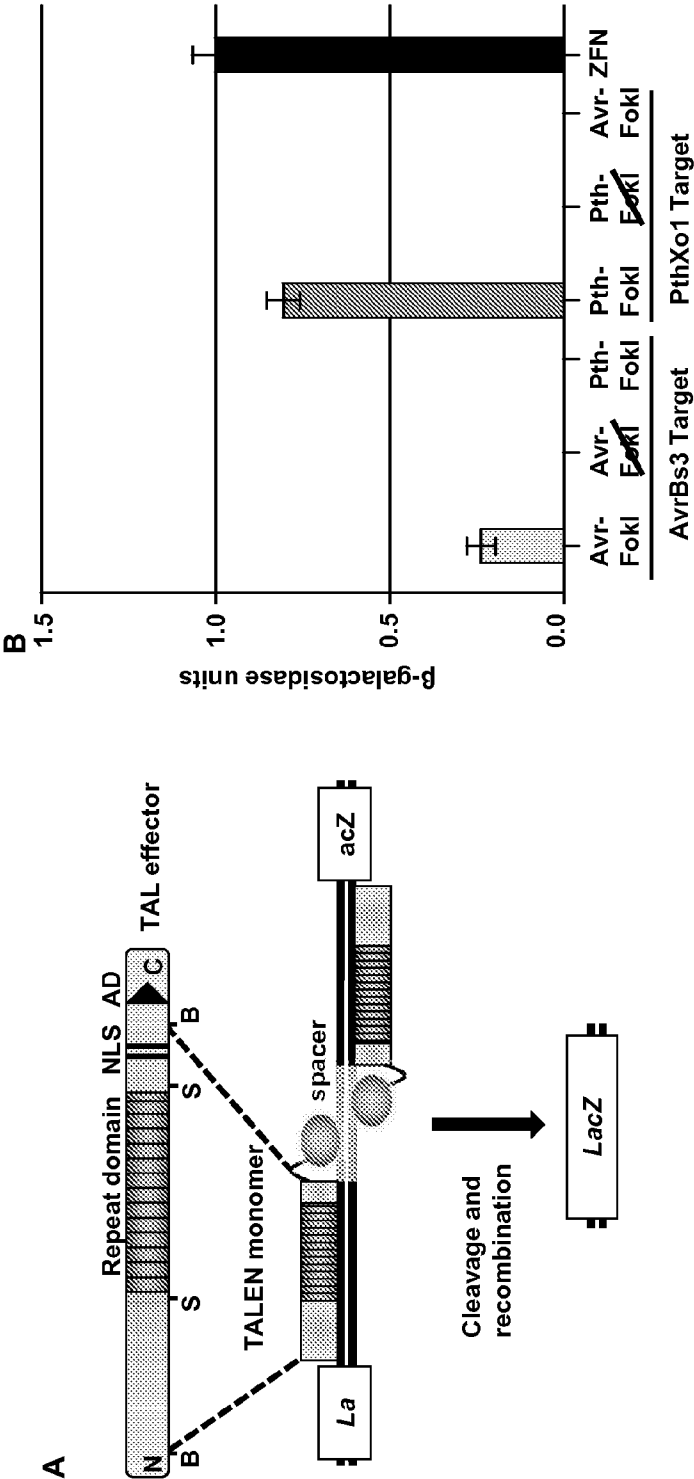
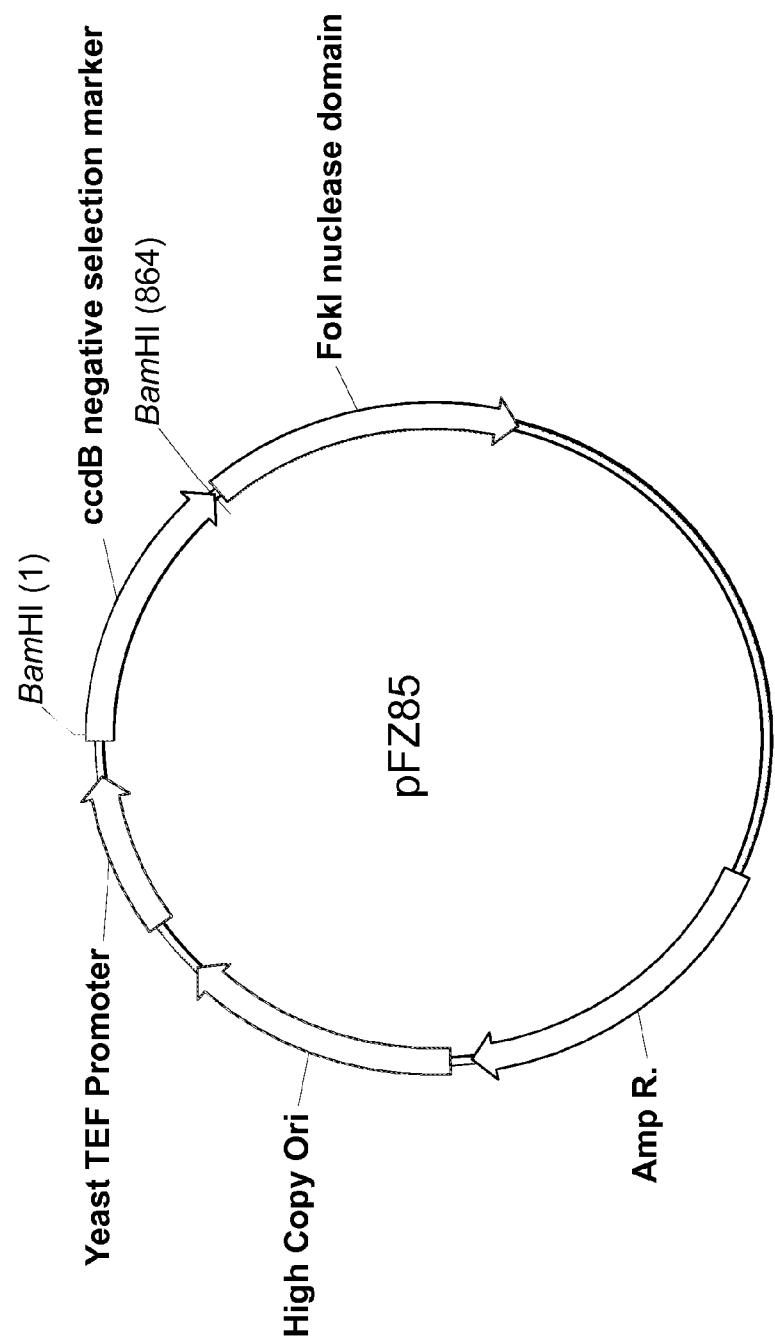


Figure 23

MDPIRSRTPSPARELLPGPQPDVRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQFDPS
LLDTSLLDSMPAVGTPHTAAAPAECEDEVQSGLRAADPPPTVRVAVTAARPPRAKPAAPRRRAAQPSDASPAAQVDLRTL
GYSQQQEKIKPKVGSTVAQHHEALVGHGFTHAHVALSRHPAALGTAVKYQDMIAALPEATHEDIVGVGKQWSGARA
LEALLTVAGELRGPPLQDGTGQLVKIAKRGGVTAVEAVHASRNALTGAPLNLTPAQVVAIASNNGGKQALETVQRLLPVLC
QAHGLTPAQVVAIASHDGGKQALETMQRLLPVLCQAHGLPPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVA
IASHGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQTH
GLTPAQVVAIASHDGGKQALETVQQLPVLCQAHGLTPDQVVAIASNIGGKQALATVQRLLPVLCQAHGLTPDQVVAIASN
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTQVQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLT
QEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG
GKQALETVQRLLPVLCQDHGLTLAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTQDQVVAIASNIGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQV
VAIASNNGKQALETVQRLLPVLCQDHGLTPNQVVAIASNNGKQALESIVAQLSRPDPALAAALTNDHLVALACLGGRPAM
DAVKKGLPHAPELIRRVNRRIGERTSHRVADYAQVVRVLEFFQCHSHPAYAFDEAMTQFGMSRNGLVQLFRVGVTELE
ARGGTLPPASQRWDRILQASGMKRAKPSPTSAQTPDQASLHAFADSLERDLDAPSPMHEGDQGTGASSRKRSDRAVT
GPSAQHSFEVRVPEQRDALHLPLSWRVKRPRTTRIGGLPDPGTPIAADLAASSTVMWEQDAAPFAGAADDFFPAFNEEEL
AWLMELLPQSGSVGGTI (SEQ ID NO:31)

[illegible]

Figure 25



[illegible]

Figure 27

pthXo1_TALEN
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPQLDGTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNNGGKQALETVQRLLPV
LCQAHGLTPAQVVAIASHDGGKQALETMQRLLPV
LCQAHGLPPDQVVAIASNIGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASHGGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPDQVVAIASHDGGKQALETVQRLLPV
CQTHGLTPAQVVAIASHDGGKQALETVQQLLPVL
CQAHGLTPDQVVAIASNIGGKQALATVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNIGGKQALETVQRLLPV
CQAHGLTPAQVVAIASHDGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
CQAHGLTQEQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPAQVVAIASHDGGKQALETVQRLLPV
CQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
CQAHGLTQEQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPAQVVAIASNIGGKQALETVQRLLPV
CQDHGLTLAQVVAIASNIGGKQALETVQRLLPV
CQAHGLTQDQVVAIASNIGGKQALETVQRLLPV
CQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
CQDHGLTLDQVVAIASNNGGKQALETVQRLLPVL
QDHGLTPDQVVAIASNNGGKQALETVQRLLPVL
QDHGLTPNQVVAIASNNGKQALESIVAQLSRPDP
ALAALTNDHLVALACLGGRPAMD VKKGLPHAPE
LIRRVNRRIGERTSHRVADYAQVVRVLEFFQCHS
HPAYAFDEAMTQFGMSRNLVQLFRRVGVTLEA
RGGTLPPASQRWDRI LQASGMKRAKPSPTSAQTP
DQASLHAFADSLERDL DAPS PMHEGDQTRASSRK
RSRSDRAVTGPSAQQAVEVRVPEQRDALHLPLSW
RVKRPRTRIWGGLPDPI SRSQLVKSELEKKSEL
RHKLKYPHEYIELIEIARNSTQDRILEMKVMEF
FMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVI
VDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHI
NPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLT
RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEE
VRRKFNNGEINF (SEO ID NO:34)

Figure 28

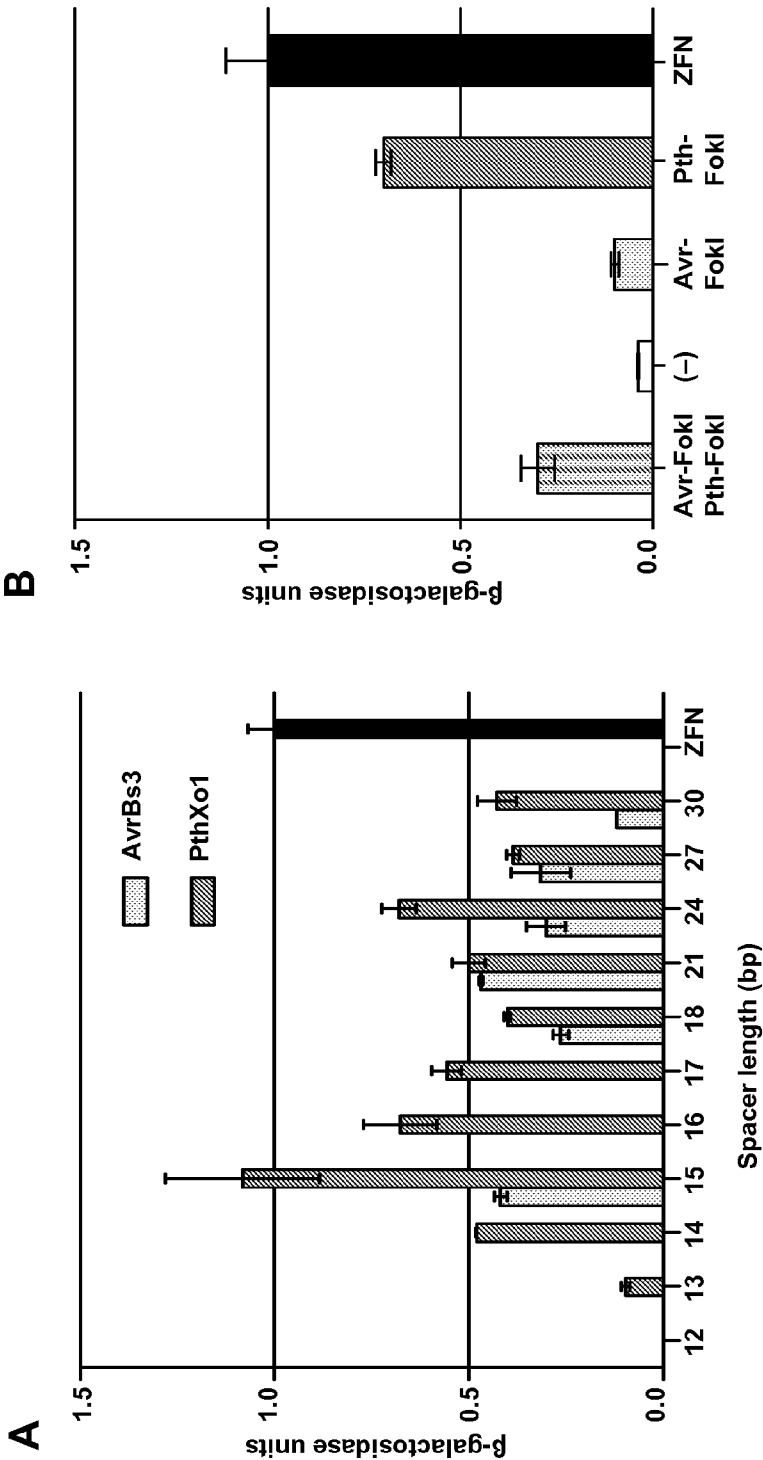


Figure 29A

Custom TALEN	1	2	3	4	5	6	7	8	9	10	11	12	13
<i>ADH1-360-12</i>	NI A	NG T	HD C	NI A	NI A	NN G	NI A	NG T	NG T	HD C	NG T	HD C	
<i>ADH1-408-12r</i>	HD C	HD C	HD C	NI A	NN G	NI A	NI A	NN G	NG T	NI A	NI A	NI A	
<i>ADH1-928-12</i>	HD C	HD C	NN G	NN G	NI A	NG T	NN G	HD C	NG T	HD C	HD C	NG T	
<i>ADH1-975-12r</i>	NI A	NN G	NI A	HD C	NI A	NI A	NI A	HD C	HD C	NI A	HD C	NI A	
<i>gridlock-2356-13r</i>	NI A	HD C	HD C	HD C	HD C	NG T	HD C	NG T	HD C	HD C	NN G	HD C	NG T

NI = A HD = C NN = G NG = T

Figure 29B

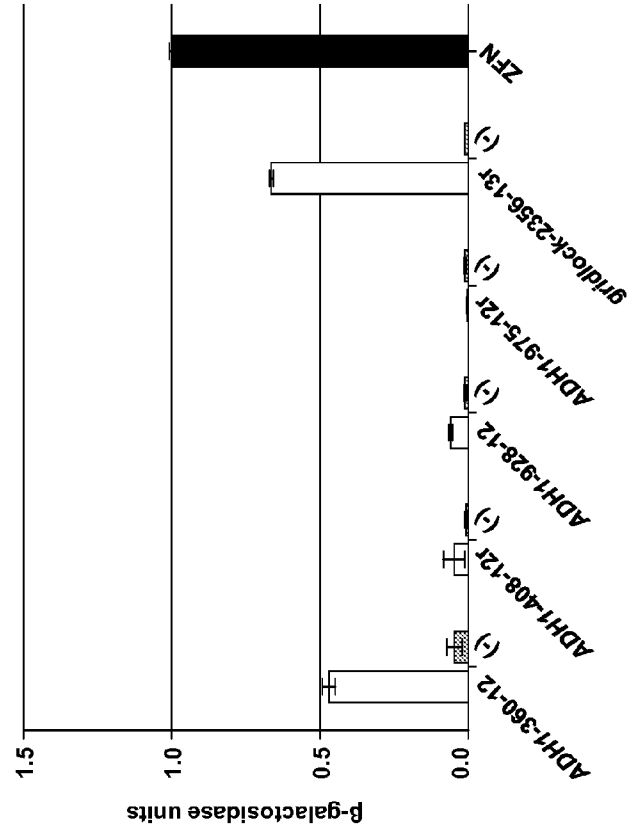


Figure 30

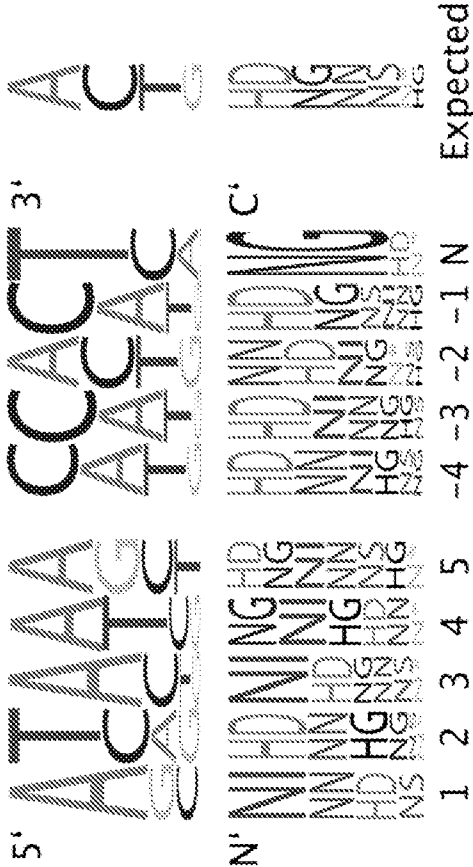


Figure 31

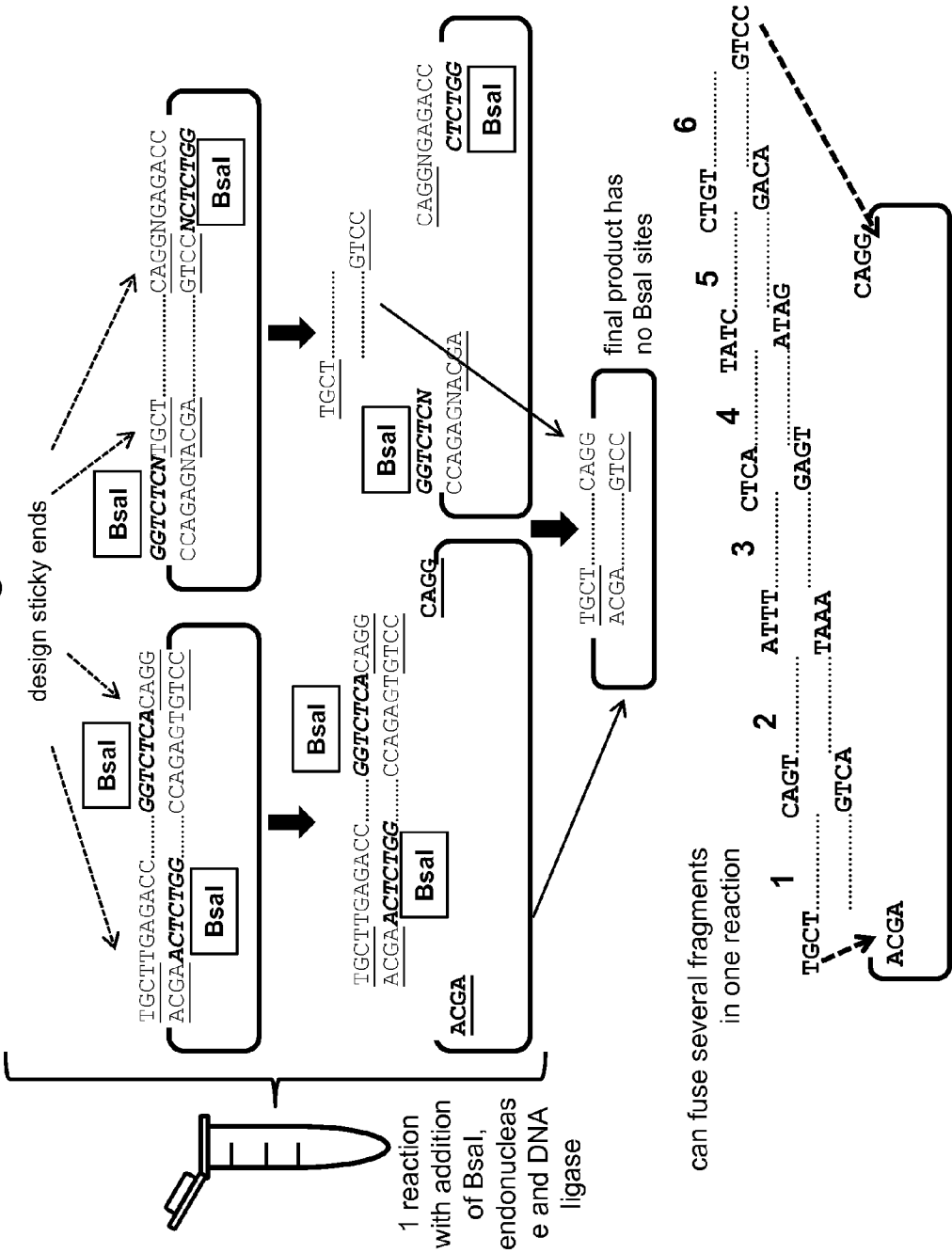


Figure 32A

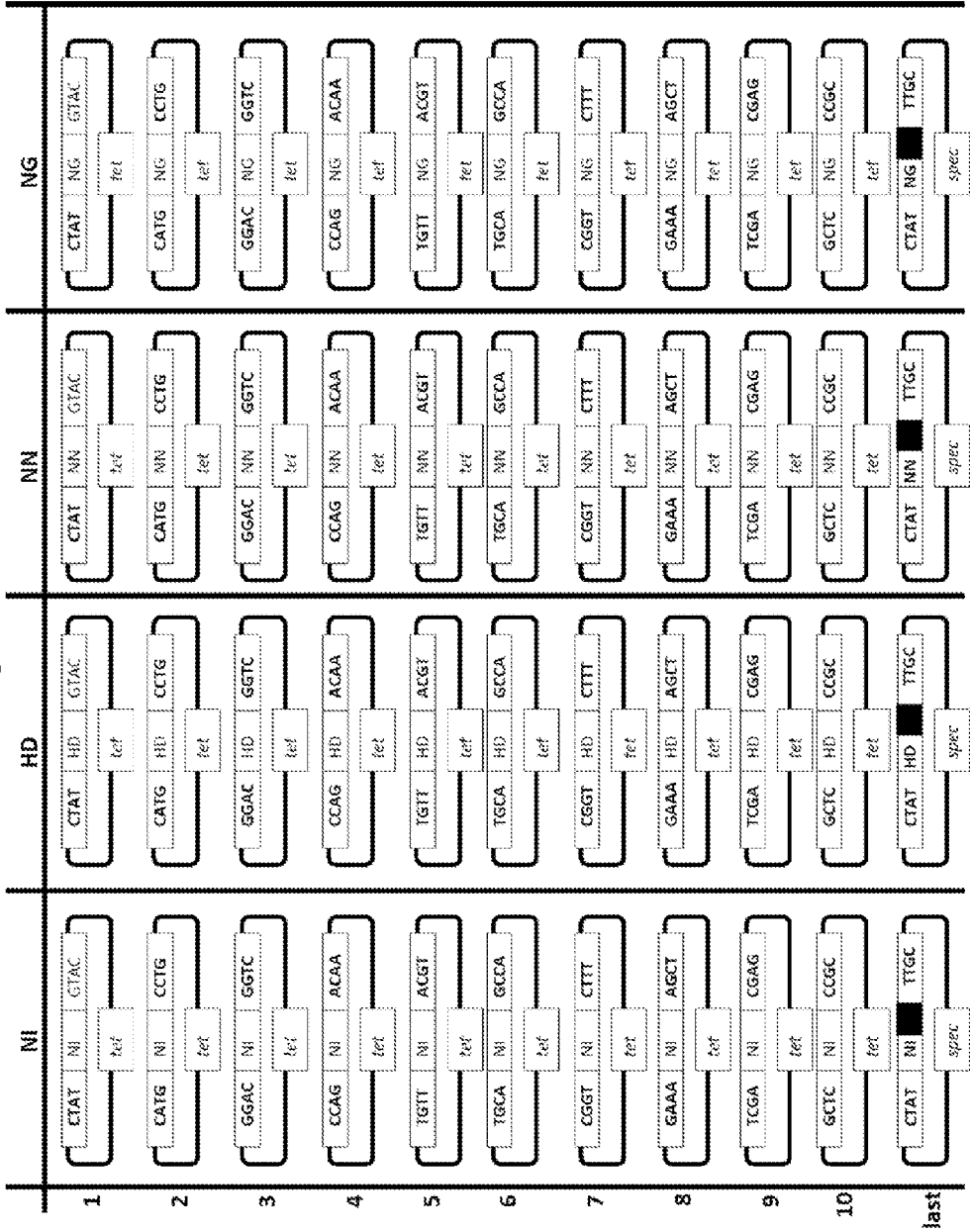


Figure 32B

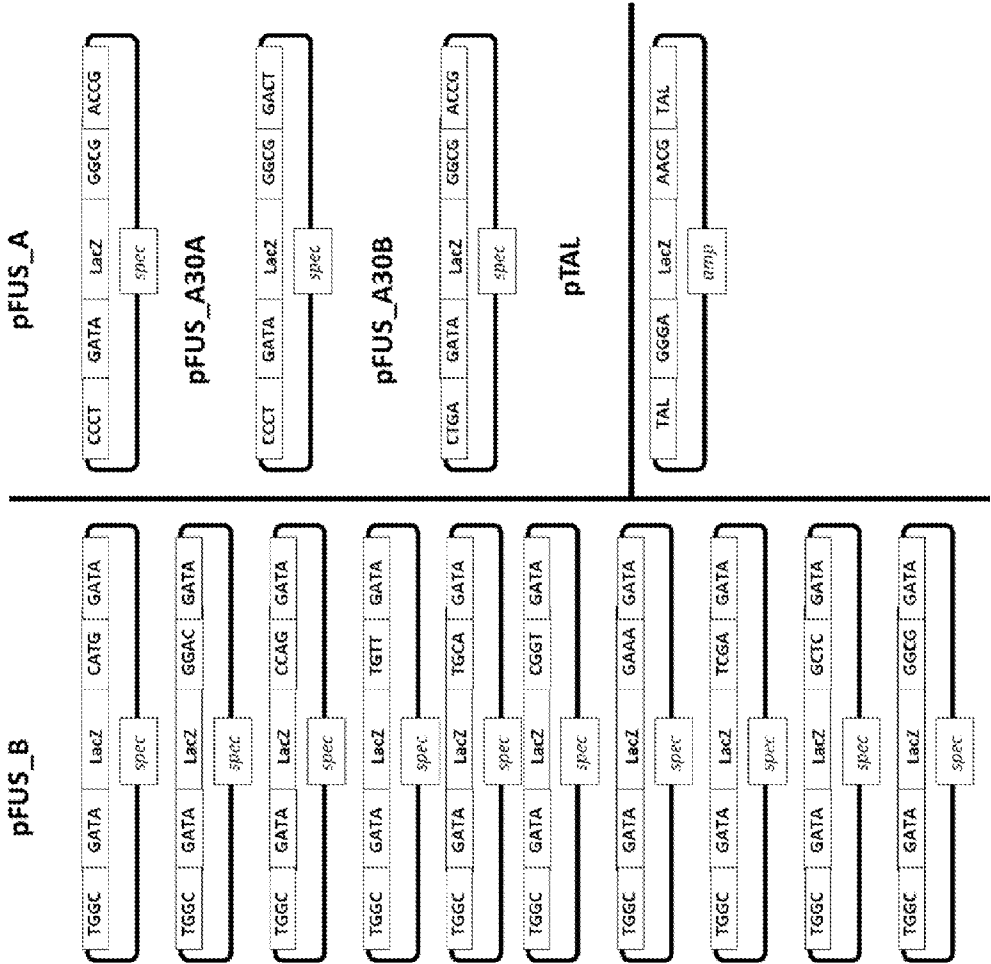


Figure 33

TALEN sequence: NI NN HD HD HD NI NI NG HD NG HD NI NI NG HD NG

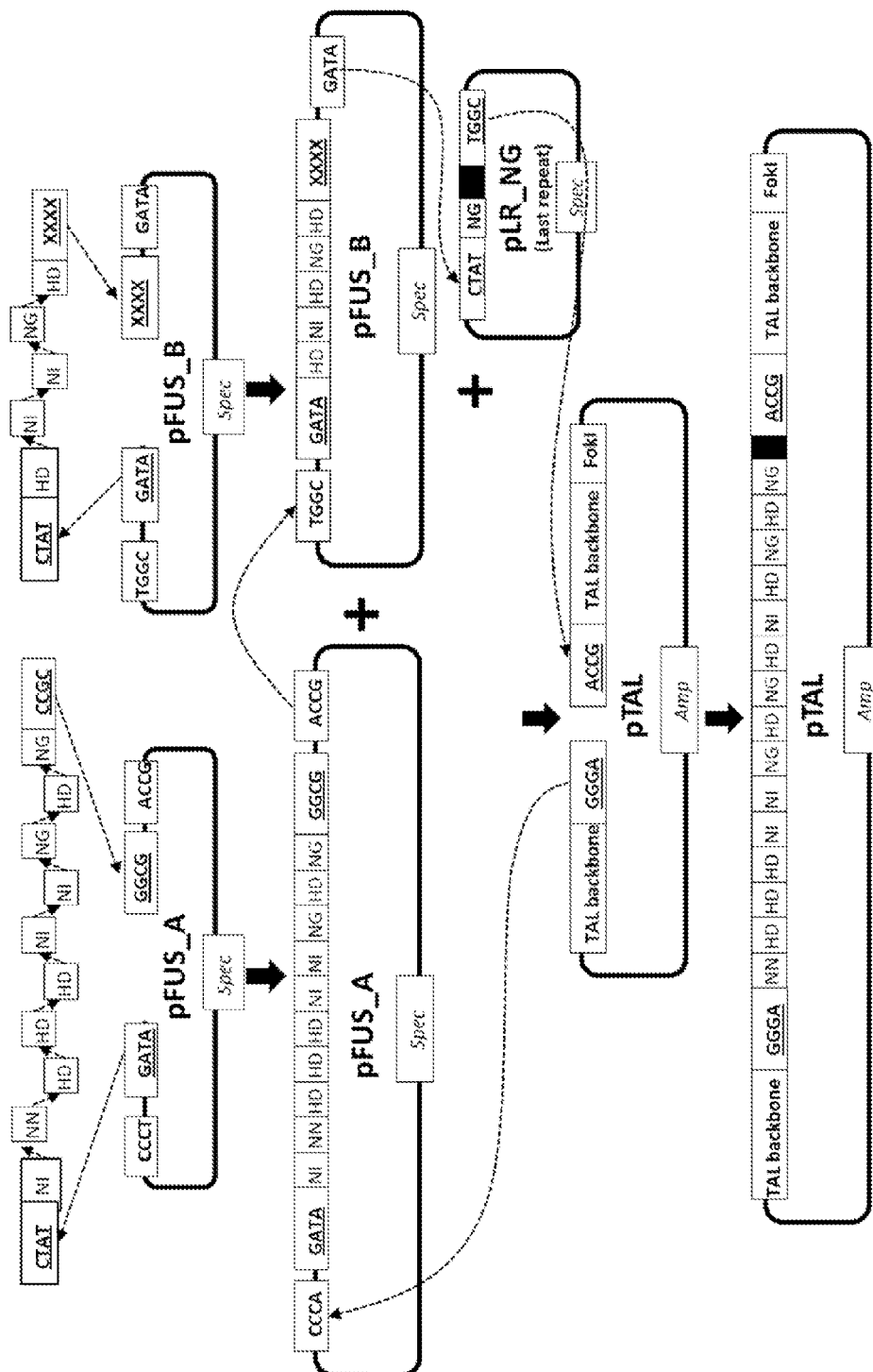


Figure 34A

telomerase-TALEN124_9RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAFAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLRLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQKEIKPKVRSSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPQLDGTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALESIQAQLSR
RDPALAAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRILQASGMKRAKPSPTSA
QTPDQTSIHAFADSLERDLDAKSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAQVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPIRSQSLVKSELEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWVKVYPSSVTEFKFLFVSGHFKGNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:35)

gridlock-TALEN105_10RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQQEKIKPKVRSVTQAHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQTS LHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAITYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEM QRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:36)

[illegible]

adh1-TALEN63_12RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDVRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPTHATAAPAWEDEAQSALR
AADDDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQQEKIKPKVRSVTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFFMKVYGYRGKHLGGSRKPDGAITYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRRKFNNGEINF (SEQ ID NO:38)

[illegible]

adh1-TALEN73_12RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLRLPFDP
SLLDTSLLDSMPAVGTPHTAAAPAWEDEAQSALR
AADDDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSVTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGLPH
APEFIRRVNRRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFLRRVGVT
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQTS LHAFADSLERDL DAPS MHEGDQTR
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYG YRGKHLGGSRKPDGAIIYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEM QRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRRKFNNGEINF (SEQ ID NO:40)

[illegible]

gridlock-TALEN106_13RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPHTAAAPAWEDEAQSALR
AADDDPPPTVRVAVTAARPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSOVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWVKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:42)

adh1-TALEN64_15RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPTHATAAAPAEWDEAQSAL
AADDDPPPVTVRVAVTAARPPRAKPAPRRRAAQPS
ASPAAQVDLRTLGYSQQQQEKIKPKVRSOVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYG YRGKHLGGSRKPDGA IYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEM QRYVEENQTRN
KHINPNEWVKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LT
LEEVR RKFNNGEINF (SEQ ID NO:43)

adh1-TALEN69_15RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLRLPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSVTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDGTQGLVKIAKRGVVTAMEAVHASRNL
TGAPLNLTPAQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:44)

[illegible]

[illegible]

telomerase-TALEN121_15RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLRLPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDDPPPTVRVAVTAARPPrAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSVTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI I
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNI GGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNI GGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQTS LHAFADSLERDL DAPS PMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SEL RHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAII YTVGSPIDY
GVIVDTKAYSGGYNLP IGQADEM QRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LT
LEEVR RKFNNGEINF (SEQ ID NO:47)

telomerase-TALEN126_15RVDs
MASSPPKKRKVSWKDASGWSRMHADPIRRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLSPSPAPSAPFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPTHATAAPAWEDEAQSAL
AADDPPTVRVAVTAAARPPRAKPAPRRRAAQPS
ASPAAQVDLRTLGYSSQQQEKIKPKVRSOVAQH
EALVGHGFTHAHIVALSQHPAALGTVAITYQHI
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGKGQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYG YRGKHLGGS RKPDGAIIYTVGSPIDY
GVIVDTKAYSGGYNLP I GQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:48)

gridlock-TALEN107_16RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPHTAAAPAWEDEAQSALR
AADDDPPPPTVRVAVTAAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQOQEIKPKVIRSTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKWQSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTA MEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKG LPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQTSLHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVRKRPRTRIWGGLPDPI SRSNQLVKSELEEK
SELRKRLKYPHEYIELIEIARNSTQDRILEMKV
MEFFFMKVYGYRGKHLGGSRKPDGAITYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEMQRVVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LT
LEEVRKFNNGEINF (SEQ ID NO:49)

[illegible]

telomerase-TALEN131_16RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLSPSPAPSAPFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPTHATAAPAWEDEAQSAL
AADDPPTVRVAVTAAARPPRAKPAPRRRAAQPS
ASPAAQVDLRTLGYSQOQQEKIKPKVRSVTQAHH
EALVGHGFTHAHIVALSOHPAALGTVAVTYQHII
TALPEATHEDIVGVGKWQSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTA MEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKG LPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLFRRVGVTE
FEARYGTLPPASQRWDRILOASGMKRAKPSPTS
QTPDQTS LHAFADSLERDL DAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRIHVWGLPDPI SRSQLVKSELEEKK
SEL RHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYG YRGKHLGSGSRKPDGAITYVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LT
LEEVRRKFNNGEINF (SEQ ID NO:51)

Figure 34R

telomerase-TALEN136_17RVDs

MASSPPKKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVIRSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDL DAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQA VEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNKYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRRKFNNGEINF (SEQ ID NDO:52)

adh1-TALEN60_18RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDVRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLDDTSLDLSMPAVGTPTHATAAPAWEDEAQSALR
AADDPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSOVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGRPALDAVKKGKLPH
APEFIRRVRNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVLQFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQTS LHAFADSLERDL DAPS PMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYG YRGKHLGGSRKPDGA IYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEM QRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:53)

Figure 34T

tt4-TALEN85_18RVDs

MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALESIVAQLSR
PDPALAAALTNDHLVALACLGGRPAMDAVKKGLPH
APELIRRVRNRRIGERTSHRVADYAQVVRVLEFFQ
CHSHPAYAFDEAMTQFGMSRNLVQLFRRVGVTE
LEARGGTLPPASQRWDRILQASGMKRAKPSPTSA
QTPDQASLHAFADSLERDL DAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPIRSQLVKSELEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:54)

gridlock-TALEN102_18RVDs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFD
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSQQQQEKIKPKVRSVTVAQH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
TALPEATHEDIVGVGKWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALESIVAQLSR
PDPALAALTNDHLVALACLGRPAMDAVKKGLPH
APELIRRVRNRIRIGERTSHRVADYAQVVRVLEFFQ
CHSHPAYAFDEAMTQFGMSRNLVLQFRRVGVTE
LEARGGTLPPASQRWDRI LQASGMKRAKPSPTS
QTPDQASLHAFADSLERDL DAPS MHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAITYTVGSPIDY
GVIVDTKAYSGGYNLPI GQADEMQRVVEENQTRN
KHINPNEWVKVYPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:55)

Figure 35

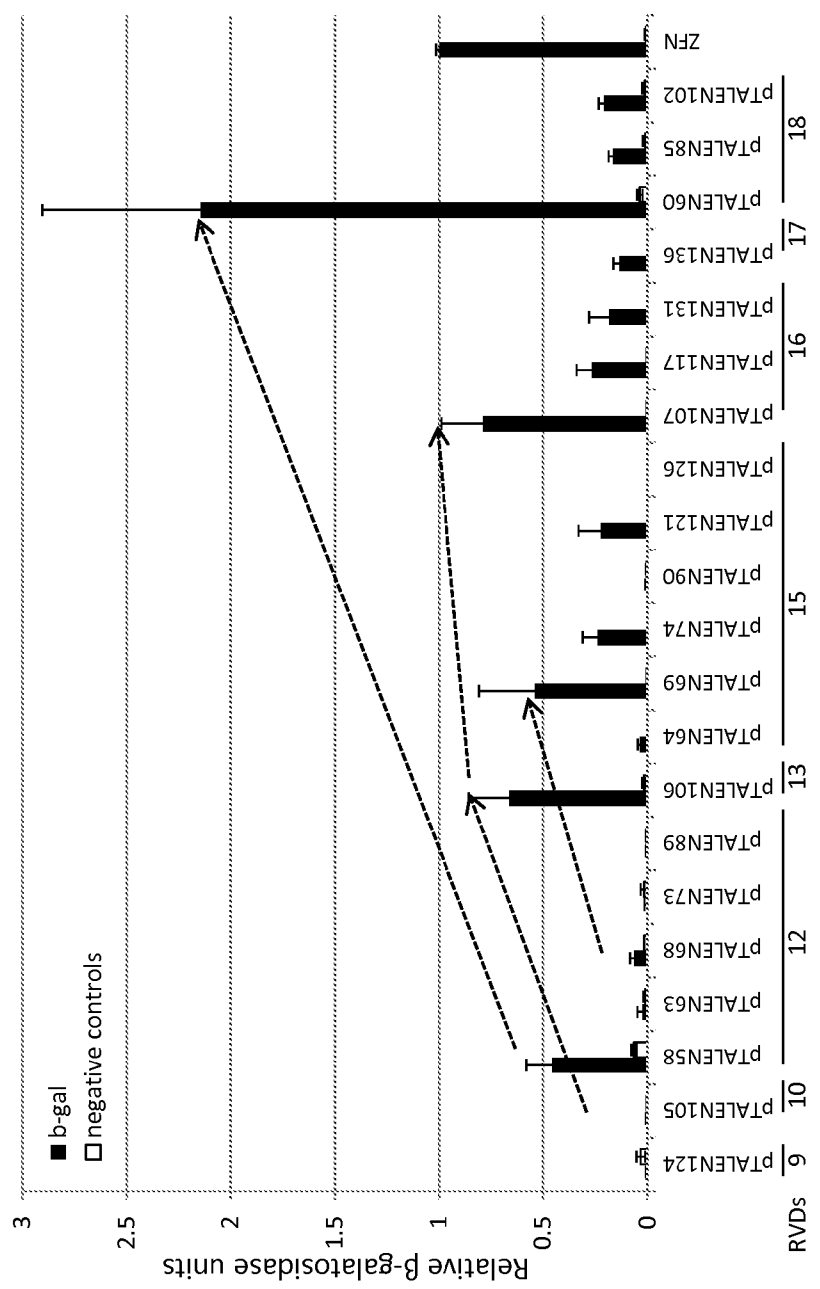


Figure 36

Adh1 target TCCGGATGCTCCTCTTGACAAAGGCTCTGATTGTCAGTTGGTTGTCTA (SEQ ID NO:57)

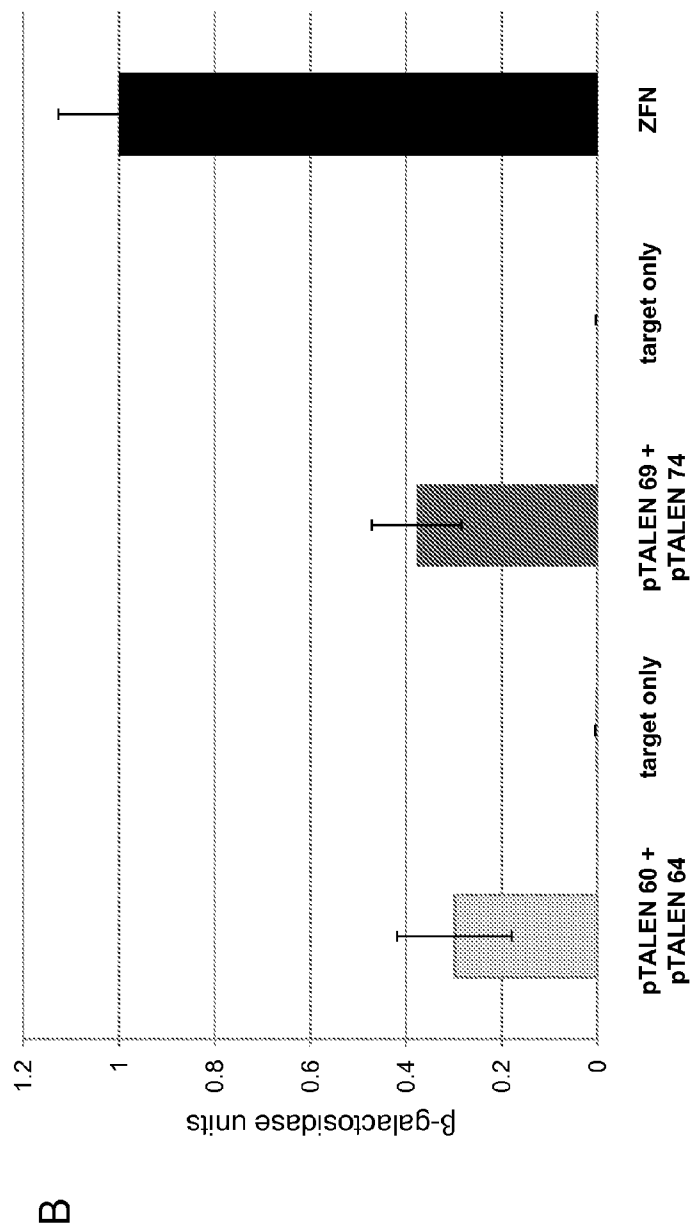
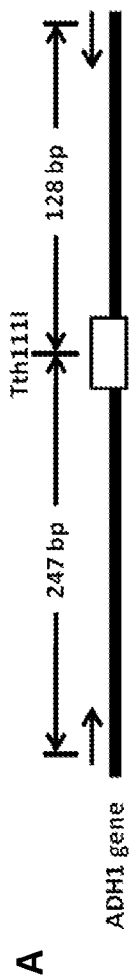


Figure 37



B

	1	10	20	30	40	48	
ADH1	CCGGATGCTCCCTCTT		GACAAAGGCTCTGTATTGTC		AGTTGTGGTTTGCT		(SEQ ID NO:58)
<u>Unmodified</u>	CCGGATGCTCCCTCTT		GACAA-----TTGTC		AGTTGTGGTTTGCT		-8 (SEQ ID NO:59)
	CCGGATGCTCCCTCTT		GACAAAG-----TATTGTC		AGTTGTGGTTTGCT		-5 (SEQ ID NO:60)
	CCGGATGCTCCCTCTT		GACAA-----TTGTGGTTTGCT				-15 (SEQ ID NO:61)
	CCGGATGCTCCCTCTT		GACAAAG-----ATTGTC		AGTTGTGGTTTGCT		-5 (SEQ ID NO:62)
	CCGGATGCTCCCTCTT		GACAAAGGCTCTGTATTGTC		AGTTGTGGTTTGCT		0 (SEQ ID NO:58)
	CCGGATGCTCCCTCTT		GACAA-----ATTGTC		AGTTGTGGTTTGCT		-7 (SEQ ID NO:63)
	CCGGATGCTCCCTCTT		GACAAAG-----TATTGTC		AGTTGTGGTTTGCT		-4 (SEQ ID NO:64)
	CCGGATGCTCCCTCTT		GACAAAGGCTCTGTATTGTC		AGTTGTGGTTTGCT		0 (SEQ ID NO:58)
	CCGGATGCTCCCTCTT		GACAAAGGCTCTGTATTGTC		AGTTGTGGTTTGCT		0 (SEQ ID NO:58)

Figure 38

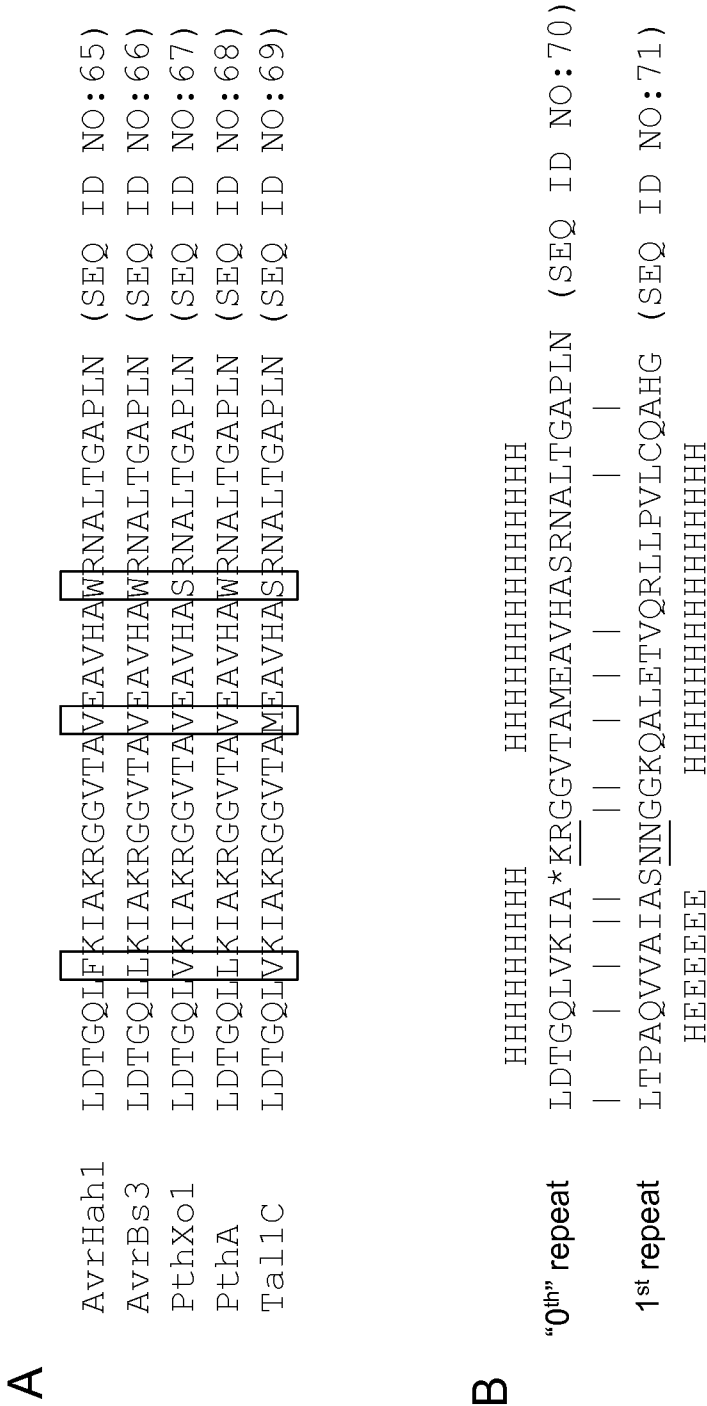
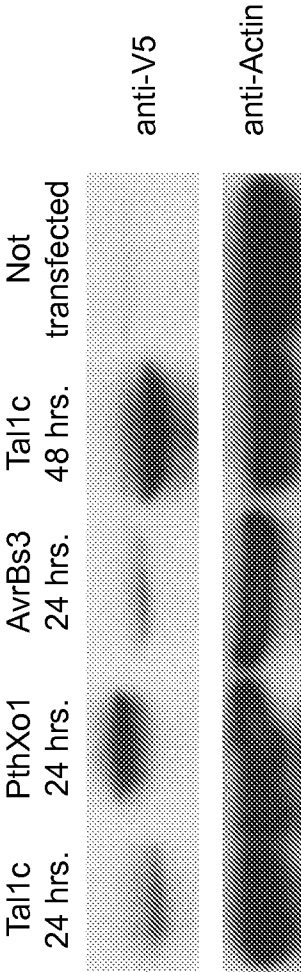


Figure 39



[illegible]

[illegible]

Figure 41A

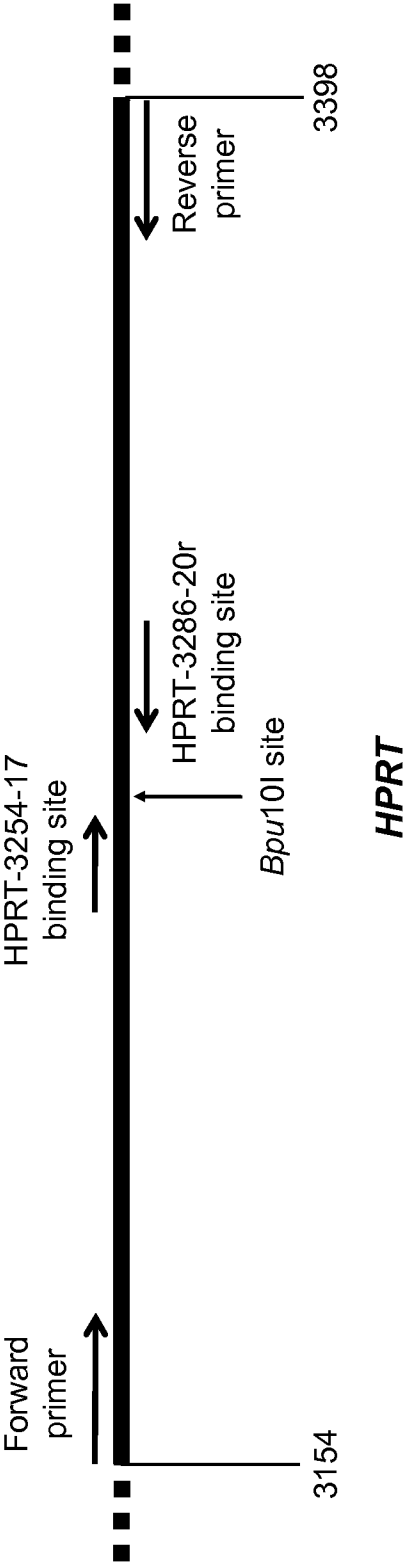
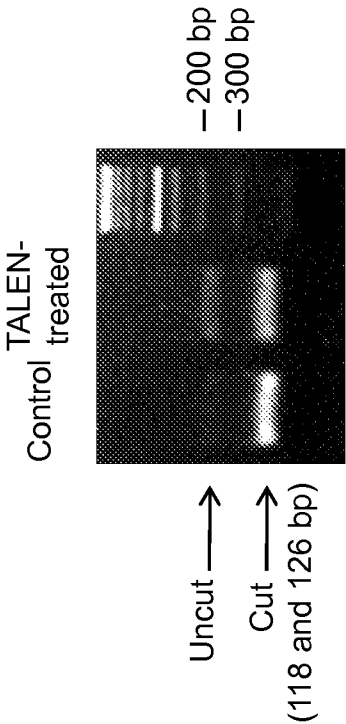


Figure 41B



TAL EFFECTOR-MEDIATED DNA MODIFICATION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. Ser. No. 12/965,590, filed on Dec. 10, 2010, which claims benefit of priority from U.S. Provisional Application Ser. No. 61/285,324, filed on Dec. 10, 2009, U.S. Provisional Application Ser. No. 61/352,108, filed on Jun. 7, 2010, and U.S. Provisional Application Ser. No. 61/366,685, filed on Jul. 22, 2010, all of which are incorporated herein by reference in their entirety.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant nos. 0820831 and 0504304, awarded by the National Science Foundation. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] The present document relates to methods for gene targeting, and particularly to methods that include the use of transcription activator-like (TAL) effector sequences.

BACKGROUND

[0004] The ability to modify chromosomes through homologous recombination (gene targeting) has been a long sought goal of biologists. In plants, for example, gene targeting may help to discern the function of plant genes, opening up new possibilities for crop improvement. For example, with gene targeting it is possible to carry out the genetic surgery required to reorchestrate metabolic pathways to create high value crops, including seed with altered oil or carbohydrate profiles, food with enhanced nutritional qualities, or plants with increased resistance to disease and stress. In animals (e.g., mammals), gene targeting may be used for treatment of disease. For example, gene targeting may be used to engineer corrections in genes that are defective due to various types of mutations. Efficient methods for such gene targeting have been difficult to achieve.

SUMMARY

[0005] TAL effectors of plant pathogenic bacteria in the genus *Xanthomonas* play important roles in disease, or trigger defense, by binding host DNA and activating effector-specific host genes (see, e.g., Gu et al. (2005) *Nature* 435:1122; Yang et al. (2006) *Proc. Natl. Acad. Sci. USA* 103:10503; Kay et al. (2007) *Science* 318:648; Sugio et al. (2007) *Proc. Natl. Acad. Sci. USA* 104:10720; and Römer et al. (2007) *Science* 318:645). Specificity depends on an effector-variable number of imperfect, typically 34 amino acid repeats (Schornack et al. (2006) *J. Plant Physiol.* 163:256). Polymorphisms are present primarily at repeat positions 12 and 13, which are referred to herein as the repeat variable-diresidue (RVD).

[0006] The present document is based in part on the fact that the RVDs of TAL effectors correspond to the nucleotides in their target sites in a direct, linear fashion, one RVD to one nucleotide, with some degeneracy and no apparent context dependence. This surprising finding represents a novel mechanism for protein-DNA recognition that enables target site prediction for new target specific TAL effector. As

described herein, these proteins may be useful in research and biotechnology as targeted chimeric nucleases that can facilitate homologous recombination in genome engineering (e.g., to add or enhance traits useful for biofuels or biorenewables in plants). These proteins also may be useful as, for example, transcription factors, and especially for therapeutic applications requiring a very high level of specificity such as therapeutics against pathogens (e.g., viruses) as non limiting examples.

[0007] In one aspect, this document features a method for modifying the genetic material of a cell, comprising (a) providing a cell containing a target DNA sequence; and (b) introducing a transcription activator-like (TAL) effector-DNA modifying enzyme into the cell, the TAL effector-DNA modifying enzyme comprising (i) a DNA modifying enzyme domain that can modify double stranded DNA, and (ii) a TAL effector domain comprising a plurality of TAL effector repeat sequences that, in combination, bind to a specific nucleotide sequence in the target DNA sequence, such that the TAL effector-DNA modifying enzyme modifies the target DNA within or adjacent to the specific nucleotide sequence in the cell or progeny thereof. The method can further comprise providing to the cell a nucleic acid comprising a sequence homologous to at least a portion of the target DNA sequence, such that homologous recombination occurs between the target DNA sequence and the nucleic acid. The cell can be a eukaryotic cell, a mammalian cell, a plant cell, or a prokaryotic cell. The target DNA can be chromosomal DNA. The introducing can comprise transfecting the cell with a vector encoding the TAL effector-DNA modifying enzyme, mechanically injecting the TAL effector-DNA modifying enzyme into the cell as a protein, delivering the TAL effector-DNA modifying enzyme into the cell as a protein by means of the bacterial type III secretion system, or introducing the TAL effector-DNA modifying enzyme into the cell as a protein by electroporation. The DNA modifying enzyme can be an endonuclease (e.g., a type II restriction endonuclease, such as FokI).

[0008] The TAL effector domain that binds to a specific nucleotide sequence within the target DNA can comprise 10 or more DNA binding repeats, and preferably 15 or more DNA binding repeats. Each DNA binding repeat can include a repeat variable-diresidue (RVD) that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T, where * represents a gap in the second position of the RVD; HG for recognizing T; H* for recognizing T, where * represents a gap in the second position of the RVD; IG for recognizing T; NK for recognizing G; HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; and YG for recognizing T. Each DNA binding repeat can comprise a RVD that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises one or more of: HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; YG for recognizing T; and NK for recognizing G, and one or more of: HD for recognizing C; NG

for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T, wherein * represents a gap in the second position of the RVD; HG for recognizing T; H* for recognizing T, wherein * represents a gap in the second position of the RVD; and IG for recognizing T.

[0009] In another aspect, this document features a method for generating a nucleic acid encoding a TAL effector specific for a selected nucleotide sequence, comprising: (1) linearizing a starter plasmid with PspXI, the starter plasmid comprising a nucleotide sequence that encodes a first TAL effector DNA binding repeat domain having a repeat variable-direidue (RVD) specific for the first nucleotide of the selected nucleotide sequence, wherein the first TAL effector DNA binding repeat domain has a unique PspXI site at its 3' end; (2) ligating into the starter plasmid PspXI site a DNA module encoding one or more TAL effector DNA binding repeat domains that have RVDs specific for the next nucleotide(s) of the selected nucleotide sequence, wherein the DNA module has XhoI sticky ends; and (3) repeating steps (1) and (2) until the nucleic acid encodes a TAL effector capable of binding to the selected nucleotide sequence. The method can further comprise, after the ligation, determining the orientation of the DNA module in the PspXI site. The method can comprise repeating steps (1) and (2) from one to 30 times.

[0010] In another aspect, this document features a method for generating a nucleic acid encoding a transcription activator-like effector endonuclease (TALEN), comprising (a) identifying a first nucleotide sequence in the genome of a cell; and (b) synthesizing a nucleic acid encoding a TALEN that comprises (i) a plurality of DNA binding repeats that, in combination, bind to the first unique nucleotide sequence, and (ii) an endonuclease that generates a double-stranded cut at a position within or adjacent to the first nucleotide sequence, wherein each DNA binding repeat comprises a RVD that determines recognition of a base pair in the target DNA, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA, and wherein the TALEN comprises one or more of the following RVDs: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T; HG for recognizing T; H* for recognizing T; IG for recognizing T; NK for recognizing G; HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; and YG for recognizing T.

[0011] The TALEN can comprises one or more of the following RVDs: HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; YG for recognizing T; and NK for recognizing G, and one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T; HG for recognizing T; H* for recognizing T; and IG for recognizing T.

[0012] The first nucleotide sequence can meet at least one of the following criteria: i) is a minimum of 15 bases long and is oriented from 5' to 3' with a T immediately preceding the site at the 5' end; ii) does not have a T in the first (5') position or an A in the second position; iii) ends in T at the last (3') position and does not have a G at the next to last position; and iv) has a base composition of 0-63% A, 11-63% C, 0-25% G, and 2-42% T.

[0013] The method can comprise identifying a first nucleotide sequence and a second nucleotide sequence in the genome of the cell, wherein the first and second nucleotide sequences meet at least one of the criteria set forth above and are separated by 15-18 bp. The endonuclease can generate a double-stranded cut between the first and second nucleotide sequences.

[0014] In another embodiment, this document features a TALEN comprising an endonuclease domain and a TAL effector DNA binding domain specific for a target DNA, wherein the DNA binding domain comprises a plurality of DNA binding repeats, each repeat comprising a RVD that determines recognition of a base pair in the target DNA, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA, and wherein the TALEN comprises one or more of the following RVDs: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T; HG for recognizing T; H* for recognizing T; IG for recognizing T; NK for recognizing G; HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; and YG for recognizing T. The TALEN can comprise one or more of the following RVDs: HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; YG for recognizing T; and NK for recognizing G, and one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T; HG for recognizing T; H* for recognizing T; and IG for recognizing T. The endonuclease domain can be from a type II restriction endonuclease (e.g., FokI).

[0015] In still another aspect, this document features a TALEN comprising an endonuclease domain and a TAL effector domain, wherein the amino acid sequence of said TALEN is selected from the group consisting of SEQ ID NO:33 to SEQ ID NO:55, SEQ ID NO:72, and SEQ ID NO:73.

[0016] This document also features a method for generating an animal, comprising: providing a eukaryotic cell comprising a target DNA sequence into which it is desired to introduce a genetic modification; generating a double-stranded cut within the target DNA sequence with a TALEN comprising an endonuclease domain and a TAL effector domain that binds to the target DNA sequence; and generating an animal from the cell or progeny thereof in which a double-stranded cut has occurred. The method can further comprise introducing into the cell an exogenous nucleic acid comprising a sequence homologous to at least a portion of the target DNA, wherein the introducing is under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the target DNA sequence in the cell or progeny thereof; and generating an animal from the cell or progeny thereof in which homologous recombination has occurred. The animal can be a mammal. The genetic modification can comprise a substitution, an insertion, or a deletion.

[0017] In yet another aspect, this document features a method for generating a plant, comprising providing a plant cell comprising a target DNA sequence into which it is desired to introduce a preselected genetic modification; generating a double-stranded cut within the target DNA sequence with a TALEN comprising an endonuclease domain and a

TAL effector domain that binds to the target DNA sequence; and generating a plant from the cell or progeny thereof in which a double-stranded cut has occurred. The method can further comprise introducing into the plant cell an exogenous nucleic acid comprising a sequence homologous to at least a portion of the target DNA sequence, wherein the introducing is under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the target DNA sequence in the cell or progeny thereof; and generating a plant from the cell or progeny thereof in which homologous recombination has occurred.

[0018] In another aspect, this document features a method for targeted genetic recombination in a cell, comprising introducing into the cell a nucleic acid encoding a TAL effector endonuclease targeted to a selected DNA target sequence; inducing expression of the TAL effector endonuclease within the cell; and identifying a cell in which the selected DNA target sequence exhibits a mutation. The mutation can be selected from the group consisting of deletion of genetic material, insertion of genetic material, and both deletion and insertion of genetic material. The method can further comprise introducing donor DNA into the cell. The cell can be an insect cell, a plant cell, a fish cell, or a mammalian cell.

[0019] In another aspect, this document features a method for generating a TAL effector having enhanced targeting capacity for a target DNA, comprising generating a nucleic acid encoding a TAL effector that comprises DNA binding domain having a plurality of DNA binding repeats, wherein each repeat comprises a RVD that determines recognition of a base pair in the target DNA, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA, wherein the generating comprises incorporating a nucleic acid encoding a variant 0th DNA binding repeat sequence with specificity for A, C, or G, thus eliminating the requirement for T at position -1 of the binding site.

[0020] In another aspect, this document features a method for generating a TAL effector having enhanced targeting capacity for a target DNA, comprising generating a nucleic acid encoding a TAL effector that comprises DNA binding domain having a plurality of DNA binding repeats, wherein each repeat comprises a RVD that determines recognition of a base pair in the target DNA, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA, wherein the generating comprises incorporating one or more nucleic acids encoding TAL effector DNA binding domains that contain RVDs having enhanced specificity for G, and wherein said RVDs are selected from the group consisting of RN, R*, NG, NH, KN, K*, NA, NT, DN, D*, NL, NM, EN, E*, NV, NC, QN, Q*, NR, NP, HN, H*, NK, NY, SN, S*, ND, NR, TN, T*, NE, NF, YN, Y*, and NQ, wherein * indicates a gap at the second position of the RVD.

[0021] This document also features a method for producing a polypeptide that selectively recognizes at least one base pair in a target DNA sequence, comprising synthesizing a polypeptide comprising a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a transcription activator-like (TAL) effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in the target DNA sequence, wherein the repeat unit is responsible for the recognition of one base pair in the DNA sequence, and wherein the hypervariable region comprises a member selected from the group consisting of: (a) HD for recognition of C/G; (b) NI for recognition of A/T; (c) NG for recognition of T/A; (d) NS for recognition

of C/G or A/T or T/A or G/C; (e) NN for recognition of G/C or A/T; (f) IG for recognition of T/A; (g) N for recognition of C/G; (h) HG for recognition of C/G or T/A; (i) H for recognition of T/A; and (j) NK for recognition of G/C. In addition, this document features a polypeptide produced by the above method, and a DNA comprising a coding sequence for the polypeptide produced by the method. Also featured is an expression cassette comprising a promoter operably linked to the above-mentioned DNA, and a non-human host cell comprising the expression cassette. In another aspect, this document features a transformed, non-human organism comprising the expression cassette.

[0022] In still another aspect, this document features a method for selectively recognizing a base pair in a DNA sequence by a polypeptide, comprising constructing a polypeptide comprising a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a TAL effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in the DNA sequence, wherein the repeat unit is responsible for the recognition of one base pair in the DNA sequence, and wherein the hypervariable region comprises a member selected from the group consisting of (a) HD for recognition of C/G; (b) NI for recognition of A/T; (c) NG for recognition of T/A; (d) NS for recognition of C/G or A/T or T/A or G/C; (e) NN for recognition of G/C or A/T; (f) IG for recognition of T/A; (g) N for recognition of C/G; (h) HG for recognition of C/G or T/A; (i) H for recognition of T/A; and (j) NK for recognition of G/C.

[0023] This document also features a method of modulating expression of a target gene in a cell, wherein cells are provided which contain a polypeptide wherein the polypeptide comprises a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a TAL effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in a DNA sequence, wherein the repeat unit is responsible for the recognition of one base pair in the DNA sequence, and wherein the hypervariable region comprises a member selected from the group consisting of (a) HD for recognition of C/G; (b) NI for recognition of A/T; (c) NG for recognition of T/A; (d) NS for recognition of C/G or A/T or T/A or G/C; (e) NN for recognition of G/C or A/T; (f) IG for recognition of T/A; (g) N for recognition of C/G; (h) HG for recognition of C/G or T/A; (i) H for recognition of T/A; and (j) NK for recognition of G/C.

[0024] In another aspect, this document features a polypeptide comprising a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a TAL effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in a DNA sequence, wherein the repeat unit is responsible for the recognition of one base pair in the DNA sequence, and wherein the hypervariable region comprises a member selected from the group consisting of (a) HD for recognition of C/G; (b) NI for recognition of A/T; (c) NG for recognition of T/A; (d) NS for recognition of C/G or A/T or T/A or G/C; (e) NN for recognition of G/C or A/T; (f) IG for recognition of T/A; (g) N for recognition of C/G; (h) HG for recognition of C/G or T/A; (i) H for recognition of T/A; and (j) NK for recognition of G/C. This document also features a DNA comprising a coding sequence for the above-mentioned polypeptide.

[0025] In another aspect, this document features a DNA which is modified to include a base pair located in a target

DNA sequence so that the base pair can be specifically recognized by a polypeptide comprising a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a TAL effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in the DNA sequence, wherein the repeat unit is responsible for the recognition of one base pair in the DNA sequence, and wherein, to receive a selective and determined recognition by the hypervariable region, the base pair is selected from the group consisting of (a) C/G for recognition by HD; (b) A/T for recognition by NI; (c) T/A for recognition by NG; (d) CT or A/T or T/A or G/C for recognition by NS; (e) G/C or A/T for recognition by NN; (f) T/A for recognition by IG; (g) C/G or T/A for recognition by N; (h) T/A for recognition by HG; (i) T/A for recognition by H; and (j) G/C for recognition by NK. Also featured are a vector comprising the above-mentioned DNA, a non-human host cell comprising the DNA, and a transformed, non-human organism comprising the DNA.

[0026] In yet another aspect, this document features a method for producing a DNA comprising a target DNA sequence that is selectively recognized by a polypeptide comprising a repeat domain, wherein the repeat domain comprises at least one repeat unit derived from a TAL effector, wherein the repeat unit comprises a hypervariable region which determines recognition of a base pair in the target DNA sequence, and wherein the repeat unit is responsible for the recognition of one base pair in the target DNA sequence, the method comprising synthesizing a DNA comprising a base pair that is capable of being recognized by the repeat unit, wherein the base pair is selected from the group consisting of (a) C/G for recognition by HD; (b) A/T for recognition by NI; (c) T/A for recognition by NG; (d) CT or A/T or T/A or G/C for recognition by NS; (e) G/C or A/T for recognition by NN; (f) T/A for recognition by IG; (g) C/G or T/A for recognition by N; (h) T/A for recognition by HG; (i) T/A for recognition by H; and (j) G/C for recognition by NK.

[0027] In one aspect, the present document features a method for modifying the genetic material of a plant cell. The method can include (a) introducing into the plant cell (i) a first recombinant nucleic acid comprising a modified target nucleotide sequence, wherein the modified target nucleotide sequence comprises one or more modifications in nucleotide sequence with respect to a corresponding target nucleotide sequence present in the plant cell, and wherein the target nucleotide sequence further comprises a recognition site for a sequence-specific TAL effector endonuclease (TALEN); and (ii) a second recombinant nucleic acid comprising a nucleotide sequence encoding the sequence-specific transcription activator-like (TAL) effector endonuclease; (b) generating a plant containing the plant cell; (c) analyzing cells, seed, or tissue obtained from the plant, or progeny thereof, for recombination at the target nucleotide sequence. The method can further include introducing into the plant cell (iii) a third recombinant nucleic acid comprising a nucleotide sequence encoding a selectable marker; and determining if the plant or progeny thereof expresses the selectable marker. The method can further include the step of screening the plant or progeny thereof for the absence of the selectable marker. The nucleotide sequence encoding the selectable marker may or may not be flanked on one or both sides by a sequence that is similar or identical to a sequence that is endogenous to the plant cell (e.g., a sequence at the site of cleavage for a second sequence-specific nuclease). The nucleotide sequence encod-

ing the selectable marker may be flanked on both sides by recognition sites for a sequence-specific recombinase. The method can further include the step of out-crossing the plant, with or without the step of screening progeny of the out-cross for the absence of the selectable marker. The first and second recombinant nucleic acids can be simultaneously introduced into the plant cell. One or both of the recombinant nucleic acids can be linearized prior to the introducing step. The first and second recombinant nucleic acids may be present in the same construct.

[0028] In another aspect, the present document features another method for modifying the genetic material of a cell. The method can include providing a primary cell containing chromosomal target DNA sequence in which it is desired to have homologous recombination occur; providing a TALEN comprising an endonuclease domain that can cleave double stranded DNA, and a TAL effector domain comprising a plurality of TAL effector repeat sequences that, in combination, bind to a specific nucleotide sequence within the target DNA in the cell; and contacting the target DNA sequence with the TALEN in the cell such that the TALEN cleaves both strands of a nucleotide sequence within or adjacent to the target DNA sequence in the cell. The method can further include providing a nucleic acid comprising a sequence homologous to at least a portion of the target DNA, such that homologous recombination occurs between the target DNA sequence and the nucleic acid. The target DNA sequence can be endogenous to the cell. The cell can be a plant cell, a mammalian cell, a fish cell, an insect cell or cell lines derived from these organisms for in vitro cultures or primary cells taken directly from living tissue and established for in vitro culture. The contacting can include transfecting the cell with a vector comprising a TALEN coding sequence, and expressing the TALEN protein in the cell, mechanically injecting a TALEN protein into the cell, delivering a TAL effector endonuclease protein into the cell by means of the bacterial type III secretion system, or introducing a TALEN protein into the cell by electroporation. The endonuclease domain can be from a type II restriction endonuclease (e.g., FokI). The TAL effector domain that binds to a specific nucleotide sequence within the target DNA can include 10 or more DNA binding repeats, more preferably 15 or more DNA binding repeats. The cell can be from any prokaryotic or eukaryotic organism.

[0029] In another aspect, the present document features a method for designing a sequence specific TALEN capable of cleaving DNA at a specific location. The method can include identifying a first unique endogenous chromosomal nucleotide sequence adjacent to a second nucleotide sequence at which it is desired to introduce a double-stranded cut; and designing a sequence specific TALEN comprising (a) a plurality of DNA binding repeat domains that, in combination, bind to the first unique endogenous chromosomal nucleotide sequence, and (b) an endonuclease that generates a double-stranded cut at the second nucleotide sequence.

[0030] The present document also features a TALEN comprising an endonuclease domain and a TAL effector DNA binding domain specific for a particular DNA sequence. The TALEN can further include a purification tag. The endonuclease domain can be from a type II restriction endonuclease (e.g., FokI).

[0031] In another aspect, the present document features a method for generating a genetically modified animal into which a desired nucleic acid has been introduced. The method can include providing a primary cell comprising an endog-

enous chromosomal target DNA sequence into which it is desired to introduce the nucleic acid; generating a double-stranded cut within the endogenous chromosomal target DNA sequence with a TALEN comprising an endonuclease domain and a TAL effector domain that binds to the endogenous chromosomal target DNA sequence; introducing an exogenous nucleic acid comprising a sequence homologous to at least a portion of the endogenous chromosomal target DNA into the primary cell under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the endogenous chromosomal target DNA; and generating an animal from the primary cell in which homologous recombination has occurred. The animal can be a mammal. The homologous sequence can be a nucleotide sequence selected from the group consisting of a nucleotide sequence that disrupts a gene after homologous recombination, a nucleotide sequence that replaces a gene after homologous recombination, a nucleotide sequence that introduces a point mutation into a gene after homologous recombination, and a nucleotide sequence that introduces a regulatory site after homologous recombination.

[0032] In still another aspect, the present document features a method for generating a genetically modified plant in which a desired nucleic acid has been introduced. The method can include providing a plant cell comprising an endogenous target DNA sequence into which it is desired to introduce the nucleic acid; generating a double-stranded cut within the endogenous target DNA sequence with a TALEN comprising an endonuclease domain and a TAL effector domain that binds to the endogenous target nucleotide sequence; introducing an exogenous nucleic acid comprising a sequence homologous to at least a portion of the endogenous target DNA into the plant cell under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the endogenous target DNA; and generating a plant from the plant cell in which homologous recombination has occurred.

[0033] In another aspect, the present document features a method for targeted genetic recombination in a cell. The method can include introducing into the cell a nucleic acid molecule encoding a TALEN targeted to a selected DNA target sequence; inducing expression of the TALEN within the cell; and identifying a cell in which the selected DNA target sequence exhibits a mutation. The mutation can be selected from the group consisting of a deletion of genetic material, an insertion of genetic material, and both a deletion and an insertion of genetic material. The method can further include introducing donor DNA into the cell. The cell can be an insect cell, a plant cell, a fish cell, or a mammalian cell.

[0034] In yet another aspect, the present document features a method for generating a nucleic acid encoding a sequence specific TALEN, comprising (1) selecting a starter plasmid comprising a nucleotide sequence that encodes a first TAL effector DNA binding repeat domain having a RVD specific for the first nucleotide of a selected nucleotide sequence, wherein the first TAL effector DNA binding repeat domain has a unique PspXI site at its 3' end; (2) linearizing the starter plasmid with PspXI; (3) ligating into the PspXI site a DNA module encoding one or more TAL effector DNA binding repeat domains that have RVDs specific for the next nucleotide(s) of the selected nucleotide sequence, wherein the DNA module has XhoI sticky ends; and (4) repeating steps (2) and (3) until the nucleic acid encodes a TALEN capable of binding to the selected nucleotide sequence. In some cases,

the method can further include, after the ligating in step (3), checking the orientation of the DNA module in the PspXI site.

[0035] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0036] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0037] FIGS. 1A-1D depict the TAL effector-DNA recognition cipher. FIG. 1A is a diagram of a generic TAL effector, showing the repeat region (open boxes) and a representative repeat sequence (SEQ ID NO:1) with the RVD underlined. FIG. 1B is a diagram showing best pattern matches (low entropy alignments) for various TAL effector RVDs and target gene promoter sequences (SEQ ID NOS:2-11). An asterisk indicates a deletion at residue 13. FIG. 1C is a diagram showing RVD-nucleotide associations in the alignments in B, plus ten more alignments obtained by scanning all rice promoters with 40 additional *X. oryzae* TAL effectors, retaining for each effector the best alignment for which the downstream gene was activated during infection. FIG. 1D is a diagram showing flanking nucleotide frequencies for the 20 TAL effector target sites. Positions are relative to the 5' end of the target site; N, length of target site. Logos were generated using WebLogo.

[0038] FIGS. 2A and 2B provide evidence that OsHen1 is activated by Tal1c of *Xanthomonas oryzae* pv. *oryzicola* strain BLS256. FIG. 2A is a picture of semi quantitative RT-PCR results, showing relative transcript abundance of OsHen1, with an actin gene for reference, in rice leaves 24 hours after inoculation with BLS256 marker exchange mutant M51, M51 carrying the empty cosmid vector (ev), M51 carrying cosmid pIJF92, which contains tal1a, tal1b, and tal1c, and the wild type (WT) strain. FIG. 2B is a schematic based on mapping of the single marker exchange mutation in M51 by rescue and end sequencing of a marker-containing XmaI fragment. The genome region, the coordinates of the rescued fragment, and the coordinates of the BLS256 genomic fragment contained in cosmid pIJF92 are shown.

[0039] FIG. 3 is a reference AvrBs3 amino acid sequence (SEQ ID NO:12).

[0040] FIG. 4 is a reference AvrBs3 nucleic acid sequence (SEQ ID NO:13).

[0041] FIG. 5 is a map of a TAL nuclease expression vector.

[0042] FIG. 6 is a map of a target reporter plasmid.

[0043] FIG. 7 is a diagram of the schematic architecture of TAL nucleases. The recognition sites of TAL DNA binding domain are represented as capital letters, while the spacer sequence is indicated in lowercase.

[0044] FIG. 8 is the amino acid sequence (SEQ ID NO:16) of the 17 and a half tandem repeats of the AvrBs3 recognition domain. Hypervariable amino acids at positions 12 and 13 are boxed.

[0045] FIG. 9 is a diagram showing a scheme for a yeast assay to test TAL effectiveness.

[0046] FIG. 10 is a graph plotting yeast assay results of AvrBs3 TAL nuclease.

[0047] FIG. 11 is a diagram showing a schematic representation of single, double, or triple AvrBs3 repeat modules and a cloning vector.

[0048] FIGS. 12A and 12B depict a single representative TAL effector repeat (FIG. 12A), as well as a representative truncated repeat (FIG. 12B) that is present at the end of the repeat region in most TAL effectors. Nucleotide and encoded amino acid sequences as shown. Ns represent nucleotides encoding the RVDs, which are indicated as "XX." Numbers are given for the amino acid positions. Sequences are taken from tal1c.

[0049] FIG. 13 is a schematic depicting the tal1c gene and the process by which the repeat region was reduced to a single, truncated repeat, resulting in pCS487, also shown. M, MscI site; S, SphI site.

[0050] FIG. 14 is a schematic depicting introduction of a translationally silent mutation at the end of the original truncated repeat in pCS487 to create a PspXI and XhoI site, yielding pCS489. Sequences of codons 18-21 in the original repeat (SEQ ID NO:6) and the mutated repeat (SEQ ID NO:8) are shown. The encoded amino acid sequence (SEQ ID NO:7) was not changed by the mutation. The mutated nucleotides are italicized.

[0051] FIG. 15 is a map of pCS488, which is a kanamycin resistant plasmid encoding only the N- and C-terminal portions of tal1c, without the repeat region, in the Gateway entry vector pENTR-D (Invitrogen, Carlsbad, Calif.).

[0052] FIG. 16 is a map of the single repeat starter plasmid designated pCS493, which encodes a repeat having the RVD NI. Three other plasmids, designated pCS494, pCS495, and pCS496, were identical except for the RVDs they encode (given at right).

[0053] FIG. 17A depicts nucleotide and encoded amino acid sequences for a single repeat module with the RVD NI. The 5' XhoI compatible cohesive end, the MscI site, and the 3' PspXI/XhoI compatible cohesive end are underlined. The RVD and the nucleotides encoding it are in bold type. Three other repeat modules were constructed that are identical to that shown except for the RVD encoding sequences, which encode HD, NI, and NG, respectively. FIG. 17B is a map of the single repeat module plasmid designated pCS502, which contains the repeat encoding sequence shown in FIG. 17A. Plasmids designated pCS503, pCS504, and pCS505 also were generated, and are identical to pCS502 except for the RVDs they encode (given at right).

[0054] FIG. 18A depicts nucleotide and encoded amino acid sequences for a single repeat module with RVD NI, in which nucleotide substitutions (italicized) prevent reconstitution of the XhoI site at the 5' end following ligation into a PspXI/XhoI site and destroy the internal MscI site. The RVD and its encoding nucleotides are in bold type. Three additional repeat modules were constructed that are identical to that shown except for the RVD encoding sequences, which encode HD, NI, and NG, respectively. FIG. 18B is a schematic of a three repeat module assembled by sequentially ligating additional repeat modules into a single repeat module

plasmid. The MscI site in the first repeat and the PspXI site at the 3' end remain unique, and the entire module is flanked by two XhoI sites.

[0055] FIG. 19 is a list of the complete set of one-, two-, and three-repeat module plasmids.

[0056] FIG. 20 is a flow chart depicting the steps in a method that can be used to assemble any sequence of repeats into the tal1c "backbone" to generate a custom TAL effector gene.

[0057] FIGS. 21A and 21B are schematics depicting assembly of repeat modules in construction of TAL endonucleases that will target the nucleotide sequences shown. In FIG. 21A, repeat modules from plasmids designated pCS519, pCS524, pCS537, pCS551, pCS583, and pCS529 are sequentially added to the sequence in the starter plasmid designated pCS493, resulting in plasmids designated pMAT55, pMAT56, pMAT57, pMAT58, pMAT59, and pMAT60. In FIG. 21B, repeat modules from plasmids designated pCS530, pCS533, pCS522, and pCS541 are sequentially added to the sequence in the plasmid designated pMAT1, resulting in plasmids designated pMAT61, pMAT62, pMAT63, and pMAT64.

[0058] FIG. 22A is a schematic of a TAL effector protein. BamHI fragments (denoted by B's) were fused to the catalytic domain of the FokI endonuclease to create TALENs. N, N-terminus; NLS, nuclear localization signal; B, BamHI site, AD, acidic activation domain. FIG. 22B is a graph plotting activity of TALENs constructed with TAL effectors AvrBs3 and PthXo1. Avr-FokI, AvrBs3 TALEN; Pth-FokI, PthXo1 TALEN, Avr-FokI and Pth-FokI, AvrBs3 and PthXo1 fusions to a catalytically inactive version of FokI (Bitinaite et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:10570-10575); ZFN, zinc finger nuclease containing the Zif268 DNA binding domain (Porteus and Baltimore (2003) *Science* 300:763).

[0059] FIG. 23 is a reference PthXo1 amino acid sequence (SEQ ID NO:31).

[0060] FIG. 24 is a reference PthXo1 nucleic acid sequence (SEQ ID NO:32).

[0061] FIG. 25 is a diagram of the pFZ85 vector.

[0062] FIG. 26 shows the amino acid sequence of avrBs3_TALEN (SEQ ID NO:33).

[0063] FIG. 27 shows the amino acid sequence of pthXo1_TALEN (SEQ ID NO:34).

[0064] FIG. 28A is a graph plotting activity of AvrBs3 and PthXo1 TALENs on targets with different spacer lengths. ZFN, Zif268-derived zinc finger nuclease. FIG. 28B is a graph plotting activity of a heterodimeric TALEN. Activity in yeast containing PthXo1-FokI and AvrBs3-FokI expression vectors and a plasmid with a target consisting of recognition sites for each, in head to tail orientation separated by 15 bp is shown (Avr-FokI, Pth-FokI). Also shown for reference is activity of AvrBs3 (Avr-FokI) and PthXo1 (Pth-FokI) TALENs individually and Zif268 (ZFN) on their respective targets. As a negative control, a yeast culture with only the target site plasmid for Avr-FokI, Pth-FokI was assayed for LacZ activity (denoted as (-)).

[0065] FIG. 29A is a table showing the RVD sequences of individual custom TALENs and their respective DNA recognition sequences. FIG. 29B is a graph plotting the activity of custom TALENs. (-), negative control with target site plasmids only; ZFN, zinc finger nuclease positive control.

[0066] FIG. 30 is a depiction of the nucleotide and RVD frequencies at the termini of 20 target and TAL effector pairs.

[0067] FIG. 31 is a schematic of the Golden Gate cloning system [Engler et al. (2008) *PLoS One* 3:e3647; and Engler et al. (2009) *PLoS One* 4:e5553].

[0068] FIGS. 32A and 32B depict a set of 58 plasmids for assembly and cloning of custom TAL effector repeat encoding arrays using the Golden Gate cloning approach as described herein. Tet, tetracycline resistance gene, a marker for plasmid selection; spec, spectinomycin resistance gene, a marker for plasmid selection; amp, ampicillin resistance gene, a marker for plasmid selection.

[0069] FIG. 33 is a schematic of a method for assembly and cloning of custom TAL effector repeat encoding arrays by the Golden Gate approach using the set of plasmids shown in FIG. 32. For illustration purposes, assembly of an arbitrary repeat array is shown. spec, spectinomycin resistance gene, a marker for plasmid selection; amp, ampicillin resistance gene, a marker for plasmid selection.

[0070] FIGS. 34A-34U show the amino acid sequences of TALENs generated as described in Example 9 herein. FIG. 34A, telomerase-TALEN124; FIG. 34B, gridlock-TALEN105; FIG. 34C, adh1-TALEN58; FIG. 34D, adh1-TALEN63; FIG. 34E, adh1-TALEN68; FIG. 34F, adh1-TALEN73; FIG. 34G, adh1-TALEN89; FIG. 34H, gridlock-TALEN106; FIG. 34I, adh1-TALEN64; FIG. 34J, adh1-TALEN69; FIG. 34K, adh1-TALEN74; FIG. 34L, tt4-TALEN90; FIG. 34M, telomerase-TALEN121; FIG. 34N, telomerase-TALEN126; FIG. 34O, gridlock-TALEN107; FIG. 34P, gridlock-TALEN117; FIG. 34Q, telomerase-TALEN131; FIG. 34R, telomerase-TALEN136; FIG. 34S, adh1-TALEN60; FIG. 34 T, tt4-TALEN85; FIG. 34U, gridlock-TALEN102.

[0071] FIG. 35 is a graph plotting TALEN activity as measured by the yeast assay using custom TALEN monomers of increasing length (9-, 10-, 12-, 13-, 15-, 16-, 17-, or 18 mers). The TALENs were targeted to *Arabidopsis* and zebrafish genes, as indicated.

[0072] FIG. 36A is a diagram showing two different DNA target sequences from the *Arabidopsis* ADH1 gene that are targeted by two TALEN pairs. FIG. 36B is a graph plotting yeast assay data for functional TALEN pairs that target the *Arabidopsis* ADH1 gene.

[0073] FIG. 37A is a schematic of a restriction endonuclease assay used to detect TALEN-induced mutations in *Arabidopsis* protoplasts. FIG. 37B shows the sequences of nine clones from undigested DNA in the restriction endonuclease assay. Six of the clones have mutations introduced by non-homologous end joining (NHEJ).

[0074] FIG. 38A shows 0th repeat sequences of several phylogenetically distinct TAL effectors, AvrHah1 from *Xanthomonas gardneri*, AvrBs3 from *X. campestris* pv. *vesicatoria*, PthXo1 from *X. oryzae* pv. *oryzae*, PthA from *X. citri*, and Tal1c from *X. oryzae* pv. *oryzicola*. Polymorphic positions are boxed. FIG. 38B is a schematic showing the 0th and 1st repeats of PthXo1. The "0th" repeat immediately precedes the 1st repeat, shows 35% identity, and has a similar predicted secondary structure. The RVD of the 1st repeat and the candidate analogous residues of the 0th repeat are underlined. *, gap; H, helix; E, extended. The structure was predicted using JPred (Cole et al. (2008) *Nucl. Acids Res.* 36:W197-W201).

[0075] FIG. 39 shows a western blot of total protein isolated from human embryonic kidney 293T cells transfected with plasmids encoding VS-tagged TAL effector proteins AvrBs3, PthXo1, and Tal1c, as indicated, following immu-

nodection using a mouse-antiV5 antibody. Immunolabeled actin is shown as a control for equivalent loading in each lane.

[0076] FIG. 40A shows the amino acid sequence of TALEN HPRT-3254-17, and FIG. 40B shows the amino acid sequence of TALEN HPRT-3286-20r.

[0077] FIG. 41A is a schematic showing the TALEN-targeted site in the human chromosomal HPRT gene. Binding sites for the HPRT-3254-17 and HPRT-3286-20r TALENs, the Bpu10I site in the spacer between those sites, and the primer sites for amplification of the region are indicated. Coordinates at the bottom give distance in base pairs from the first nucleotide of the coding sequence. FIG. 41B shows the results of Bpu10I digestion of products of PCR amplification of the region shown in FIG. 41A using genomic DNA isolated from TALEN-treated and untreated cells as templates. Genomic DNA was digested with Bpu10I prior to amplification. DNA fragments were separated by agarose gel electrophoresis and visualized using ethidium bromide.

DETAILED DESCRIPTION

[0078] The present patent application provides materials and methods related to sequence specific DNA recognition mediated by TAL effectors. As described herein, the primary amino acid sequences of TAL effectors dictate the nucleotide sequences to which they bind. The inventors have found that relationships between TAL effector amino acid sequences and their DNA target sequences are direct, enabling target site prediction for TAL effectors, and also allowing for TAL effector customization to bind to particular nucleotide sequences. Such prediction and customization can be harnessed for a variety of purposes. In one example, particular TAL effector sequences can be fused to endonuclease sequences, allowing for endonuclease targeting to specific DNA sequences, and subsequent cutting of the DNA at or near the targeted sequences. Cuts (i.e., double-stranded breaks) in DNA can dramatically increase the frequency of homologous recombination. Thus, in combination with DNA constructs that carry sequences having a high degree of sequence similarity to a particular target DNA sequence, TALENs can be used to facilitate site directed mutagenesis in complex genomes, that is, to knock out or alter gene function, or to add genes or other sequences with great precision and high efficiency.

[0079] Thus, included in the subject matter provided herein are, inter alia, materials and methods for making genetically modified organisms (including, without limitation, plants, fungi, *Drosophila*, nematodes, zebrafish, mice, other mammals and humans). Such methods can include, for example, transfecting a cell with several recombinant nucleic acids. For example, a cell (e.g., a eukaryotic cell) can be transformed with a first recombinant nucleic acid construct containing a donor nucleotide sequence that includes alterations relative to a corresponding target nucleotide sequence found within the cell, and a second recombinant nucleic acid construct encoding a TAL-nuclease. In some embodiments, the cell also can be transformed with a third recombinant nucleic acid construct encoding a selectable marker. A nucleic acid sequence from the donor nucleic acid construct can become incorporated into the genome of the transformed cell as described herein. For example, plant cells produced using methods as described herein can be grown to produce plants having the altered donor nucleotide sequence incorporated into their genomes. Seeds from such plants can be used to produce plants having a phenotype such as, for example, an altered growth characteristic (e.g., increased resistance or tolerance

to various biotic and abiotic stresses), altered appearance (e.g., altered color or height), or altered composition (e.g., increased or decreased levels of carbon, nitrogen, oil, protein, carbohydrate (e.g., sugar or starch), amino acid, fatty acid, or secondary metabolites) with respect to unmodified plants.

Polynucleotides and Polypeptides

[0080] Isolated nucleic acids and polypeptides are provided herein. The terms “nucleic acid” and “polynucleotide” are used interchangeably, and refer to both RNA and DNA, including cDNA, genomic DNA, synthetic (e.g., chemically synthesized) DNA, and DNA (or RNA) containing nucleic acid analogs. Polynucleotides can have any three-dimensional structure. A nucleic acid can be double-stranded or single-stranded (i.e., a sense strand or an antisense single strand). Non-limiting examples of polynucleotides include genes, gene fragments, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers, as well as nucleic acid analogs.

[0081] The polypeptides of the present document (such as TAL effector-DNA modifying enzyme as non-limiting example) can be introduced in a cell by using a vector encoding said polypeptides for example or as polypeptides per se by using delivery vectors associated or combined with any cellular permeabilization techniques such as sonoporation or electroporation or derivatives of these techniques.

[0082] As used herein, “isolated,” when in reference to a nucleic acid, refers to a nucleic acid that is separated from other nucleic acids that are present in a genome, e.g., a plant genome, including nucleic acids that normally flank one or both sides of the nucleic acid in the genome. The term “isolated” as used herein with respect to nucleic acids also includes any non-naturally-occurring sequence, since such non-naturally-occurring sequences are not found in nature and do not have immediately contiguous sequences in a naturally-occurring genome.

[0083] An isolated nucleic acid can be, for example, a DNA molecule, provided one of the nucleic acid sequences normally found immediately flanking that DNA molecule in a naturally-occurring genome is removed or absent. Thus, an isolated nucleic acid includes, without limitation, a DNA molecule that exists as a separate molecule (e.g., a chemically synthesized nucleic acid, or a cDNA or genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences, as well as DNA that is incorporated into a vector, an autonomously replicating plasmid, a virus (e.g., a pararetrovirus, a retrovirus, lentivirus, adenovirus, or herpes virus), or the genomic DNA of a prokaryote or eukaryote. In addition, an isolated nucleic acid can include a recombinant nucleic acid such as a DNA molecule that is part of a hybrid or fusion nucleic acid. A nucleic acid existing among hundreds to millions of other nucleic acids within, for example, cDNA libraries or genomic libraries, or gel slices containing a genomic DNA restriction digest, is not to be considered an isolated nucleic acid.

[0084] A nucleic acid can be made by, for example, chemical synthesis or polymerase chain reaction (PCR). PCR refers to a procedure or technique in which target nucleic acids are amplified. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA or total cellular RNA. Various PCR methods

are described, for example, in *PCR Primer: A Laboratory Manual*, Dieffenbach and Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. Various PCR strategies also are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid.

[0085] Isolated nucleic acids also can be obtained by mutagenesis. For example, a donor nucleic acid sequence can be mutated using standard techniques, including oligonucleotide-directed mutagenesis and site-directed mutagenesis through PCR. See, *Short Protocols in Molecular Biology*, Chapter 8, Green Publishing Associates and John Wiley & Sons, edited by Ausubel et al., 1992.

[0086] The term “polypeptide” as used herein refers to a compound of two or more subunit amino acids regardless of post-translational modification (e.g., phosphorylation or glycosylation). The subunits may be linked by peptide bonds or other bonds such as, for example, ester or ether bonds. The term “amino acid” refers to either natural and/or unnatural or synthetic amino acids, including D/L optical isomers.

[0087] By “isolated” or “purified” with respect to a polypeptide it is meant that the polypeptide is separated to some extent from the cellular components with which it is normally found in nature (e.g., other polypeptides, lipids, carbohydrates, and nucleic acids). A purified polypeptide can yield a single major band on a non-reducing polyacrylamide gel. A purified polypeptide can be at least about 75% pure (e.g., at least 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% pure). Purified polypeptides can be obtained by, for example, extraction from a natural source, by chemical synthesis, or by recombinant production in a host cell or transgenic plant, and can be purified using, for example, affinity chromatography, immunoprecipitation, size exclusion chromatography, and ion exchange chromatography. The extent of purification can be measured using any appropriate method, including, without limitation, column chromatography, polyacrylamide gel electrophoresis, or high-performance liquid chromatography.

Recombinant Constructs

[0088] Recombinant nucleic acid constructs (e.g., vectors) also are provided herein. A “vector” is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. Suitable vector backbones include, for example, those routinely used in the art such as plasmids, viruses, artificial chromosomes, BACs, YACs, or PACs. The term “vector” includes cloning and expression vectors, as well as viral vectors and integrating vectors. An “expression vector” is a vector that includes one or more expression control sequences, and an “expression control sequence” is a DNA sequence that controls and regulates the transcription and/or translation of another DNA sequence. Suitable expression vectors include, without limitation, plasmids and viral vectors derived from, for example, bacteriophage, baculoviruses, tobacco mosaic virus, herpes viruses, cytomegalovirus, retroviruses, vaccinia viruses, adenoviruses, and adeno-associated viruses. Numerous vectors and expression systems are commercially available from such corporations as Novagen (Madison, Wis.), Clontech

(Palo Alto, Calif.), Stratagene (La Jolla, Calif.), and Invitrogen/Life Technologies (Carlsbad, Calif.).

[0089] The terms “regulatory region,” “control element,” and “expression control sequence” refer to nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of the transcript or polypeptide product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, promoter control elements, protein binding sequences, 5' and 3' untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, introns, and other regulatory regions that can reside within coding sequences, such as secretory signals, Nuclear Localization Sequences (NLS) and protease cleavage sites.

[0090] As used herein, “operably linked” means incorporated into a genetic construct so that expression control sequences effectively control expression of a coding sequence of interest. A coding sequence is “operably linked” and “under the control” of expression control sequences in a cell when RNA polymerase is able to transcribe the coding sequence into RNA, which if an mRNA, then can be translated into the protein encoded by the coding sequence. Thus, a regulatory region can modulate, e.g., regulate, facilitate or drive, transcription in the plant cell, plant, or plant tissue in which it is desired to express a modified target nucleic acid.

[0091] A promoter is an expression control sequence composed of a region of a DNA molecule, typically within 100 nucleotides upstream of the point at which transcription starts (generally near the initiation site for RNA polymerase II). Promoters are involved in recognition and binding of RNA polymerase and other proteins to initiate and modulate transcription. To bring a coding sequence under the control of a promoter, it typically is necessary to position the translation initiation site of the translational reading frame of the polypeptide between one and about fifty nucleotides downstream of the promoter. A promoter can, however, be positioned as much as about 5,000 nucleotides upstream of the translation start site, or about 2,000 nucleotides upstream of the transcription start site. A promoter typically comprises at least a core (basal) promoter. A promoter also may include at least one control element such as an upstream element. Such elements include upstream activation regions (UARs) and, optionally, other DNA sequences that affect transcription of a polynucleotide such as a synthetic upstream element.

[0092] The choice of promoters to be included depends upon several factors, including, but not limited to, efficiency, selectability, inducibility, desired expression level, and cell or tissue specificity. For example, tissue-, organ- and cell-specific promoters that confer transcription only or predominantly in a particular tissue, organ, and cell type, respectively, can be used. In some embodiments, promoters specific to vegetative tissues such as the stem, parenchyma, ground meristem, vascular bundle, cambium, phloem, cortex, shoot apical meristem, lateral shoot meristem, root apical meristem, lateral root meristem, leaf primordium, leaf mesophyll, or leaf epidermis can be suitable regulatory regions. In some embodiments, promoters that are essentially specific to seeds (“seed-preferential promoters”) can be useful. Seed-specific promoters can promote transcription of an operably linked nucleic acid in endosperm and cotyledon tissue during seed development. Alternatively, constitutive promoters can promote transcription of an operably linked nucleic acid in most

or all tissues of a plant, throughout plant development. Other classes of promoters include, but are not limited to, inducible promoters, such as promoters that confer transcription in response to external stimuli such as chemical agents, developmental stimuli, or environmental stimuli.

[0093] A basal promoter is the minimal sequence necessary for assembly of a transcription complex required for transcription initiation. Basal promoters frequently include a “TATA box” element that may be located between about 15 and about 35 nucleotides upstream from the site of transcription initiation. Basal promoters also may include a “CCAAT box” element (typically the sequence CCAAT) and/or a GGGCG sequence, which can be located between about 40 and about 200 nucleotides, typically about 60 to about 120 nucleotides, upstream from the transcription start site.

[0094] Non-limiting examples of promoters that can be included in the nucleic acid constructs provided herein include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1' or 2' promoters derived from T-DNA of *Agrobacterium tumefaciens*, promoters from a maize leaf-specific gene described by Busk ((1997) *Plant J.* 11:1285-1295), kn1-related genes from maize and other species, and transcription initiation regions from various plant genes such as the maize ubiquitin-1 promoter.

[0095] A 5' untranslated region (UTR) is transcribed, but is not translated, and lies between the start site of the transcript and the translation initiation codon and may include the +1 nucleotide. A 3' UTR can be positioned between the translation termination codon and the end of the transcript. UTRs can have particular functions such as increasing mRNA message stability or translation attenuation. Examples of 3' UTRs include, but are not limited to polyadenylation signals and transcription termination sequences. A polyadenylation region at the 3'-end of a coding region can also be operably linked to a coding sequence. The polyadenylation region can be derived from the natural gene, from various other plant genes, or from an *Agrobacterium* T-DNA.

[0096] The vectors provided herein also can include, for example, origins of replication, and/or scaffold attachment regions (SARs). In addition, an expression vector can include a tag sequence designed to facilitate manipulation or detection (e.g., purification or localization) of the expressed polypeptide. Tag sequences, such as green fluorescent protein (GFP), glutathione S-transferase (GST), polyhistidine, c-myc, hemagglutinin, or Flag™ tag (Kodak, New Haven, Conn.) sequences typically are expressed as a fusion with the encoded polypeptide. Such tags can be inserted anywhere within the polypeptide, including at either the carboxyl or amino terminus.

[0097] By “delivery vector” or “delivery vectors” is intended any delivery vector which can be used in the presently described methods to put into cell contact or deliver inside cells or subcellular compartments agents/chemicals and molecules (proteins or nucleic acids). It includes, but is not limited to liposomal delivery vectors, viral delivery vectors, drug delivery vectors, chemical carriers, polymeric carriers, lipoplexes, polyplexes, dendrimers, microbubbles (ultrasound contrast agents), nanoparticles, emulsions or other appropriate transfer vectors. These delivery vectors allow delivery of molecules, chemicals, macromolecules (genes, proteins), or other vectors such as plasmids, peptides developed by Diatos. In these cases, delivery vectors are molecule

carriers. By “delivery vector” or “delivery vectors” is also intended delivery methods to perform transfection.

[0098] The terms “vector” or “vectors” refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A “vector” in the present document includes, but is not limited to, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule which may consists of a chromosomal, non chromosomal, semi-synthetic or synthetic nucleic acids. Preferred vectors are those capable of autonomous replication (episomal vector) and/or expression of nucleic acids to which they are linked (expression vectors). Large numbers of suitable vectors are known to those of skill in the art and commercially available.

[0099] Viral vectors include retrovirus, adenovirus, parvovirus (e.g. adenoassociated viruses), coronavirus, negative strand RNA viruses such as orthomyxovirus (e.g., influenza virus), rhabdovirus (e.g., rabies and vesicular stomatitis virus), paramyxovirus (e.g. measles and Sendai), positive strand RNA viruses such as picornavirus and alphavirus, and double-stranded DNA viruses including adenovirus, herpesvirus (e.g., Herpes Simplex virus types 1 and 2, Epstein-Barr virus, cytomegalovirus), and poxvirus (e.g., vaccinia, fowlpox and canarypox). Other viruses include Norwalk virus, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus, and hepatitis virus, for example. Examples of retroviruses include: avian leukosis-sarcoma, mammalian C-type, B-type viruses, D type viruses, HTLV-BLV group, lentivirus, spumavirus (Coffin, J. M., *Retroviridae: The viruses and their replication*, In *Fundamental Virology*, Third Edition, B. N. Fields, et al., Eds., Lippincott-Raven Publishers, Philadelphia, 1996).

[0100] By “lentiviral vector” is meant HIV-Based lentivirus vectors that are very promising for gene delivery because of their relatively large packaging capacity, reduced immunogenicity and their ability to stably transduce with high efficiency a large range of different cell types. Lentiviral vectors are usually generated following transient transfection of three (packaging, envelope and transfer) or more plasmids into producer cells. Like HIV, lentiviral vectors enter the target cell through the interaction of viral surface glycoproteins with receptors on the cell surface. On entry, the viral RNA undergoes reverse transcription, which is mediated by the viral reverse transcriptase complex. The product of reverse transcription is a double-stranded linear viral DNA, which is the substrate for viral integration in the DNA of infected cells. Said lentiviral vectors can be “non-integrative” or “integrative”.

[0101] By “integrative lentiviral vectors (or LV)”, is meant such vectors as non limiting example, that are able to integrate the genome of a target cell.

[0102] At the opposite by “non integrative lentiviral vectors (or NILV)” is meant efficient gene delivery vectors that do not integrate the genome of a target cell through the action of the virus integrase.

[0103] One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as “expression vectors. A vector according to the present document comprises, but is not limited to, a YAC (yeast artificial chro-

mosome), a BAC (bacterial artificial), a baculovirus vector, a phage, a phagemid, a cosmid, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule which may consist of chromosomal, non chromosomal, semi-synthetic or synthetic DNA. In general, expression vectors of utility in recombinant DNA techniques are often in the form of “plasmids” which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. Large numbers of suitable vectors are known to those of skill in the art. Vectors can comprise selectable markers, for example: neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase for eukaryotic cell culture; TRP 1 for *S. cerevisiae*; tetracyclin, rifampicin or ampicillin resistance in *E. coli*. Preferably said vectors are expression vectors, wherein a sequence encoding a polypeptide of interest is placed under control of appropriate transcriptional and translational control elements to permit production or synthesis of said polypeptide. Therefore, said polynucleotide is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a promoter operatively linked to said encoding polynucleotide, a ribosome binding site, a RNA-splicing site (when genomic DNA is used), a polyadenylation site and a transcription termination site. It also can comprise an enhancer or silencer elements. Selection of the promoter will depend upon the cell in which the polypeptide is expressed. Suitable promoters include tissue specific and/or inducible promoters. Examples of inducible promoters are: eukaryotic metallothionein promoter which is induced by increased levels of heavy metals, prokaryotic lacZ promoter which is induced in response to isopropyl- β -D-thiogalacto-pyranoside (IPTG) and eukaryotic heat shock promoter which is induced by increased temperature. Examples of tissue specific promoters are skeletal muscle creatine kinase, prostate-specific antigen (PSA), α -antitrypsin protease, human surfactant (SP) A and B proteins, β -casein and acidic whey protein genes.

[0104] Inducible promoters may be induced by pathogens or stress, more preferably by stress like cold, heat, UV light, or high ionic concentrations (reviewed in Potenza et al. (2004) *In vitro Cell Dev Biol* 40:1-22). Inducible promoter may be induced by chemicals [reviewed in Moore et al. (2006); Padidam (2003); Wang et al. (2003); and Zuo and Chua (2000)].

[0105] Delivery vectors and vectors can be associated or combined with any cellular permeabilization techniques such as sonoporation or electroporation or derivatives of these techniques.

[0106] It will be understood that more than one regulatory region may be present in a recombinant polynucleotide, e.g., introns, enhancers, upstream activation regions, and inducible elements.

[0107] Recombinant nucleic acid constructs can include a polynucleotide sequence inserted into a vector suitable for transformation of cells (e.g., plant cells or animal cells). Recombinant vectors can be made using, for example, standard recombinant DNA techniques (see, e.g., Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.).

[0108] A recombinant nucleic acid sequence as described herein can integrate into the genome of a cell via illegitimate (i.e., random, non-homologous, non site-specific) recombi-

nation, or a recombinant nucleic acid sequence as described herein can be adapted to integrate into the genome of a cell via homologous recombination. Nucleic acid sequences adapted for integration via homologous recombination are flanked on both sides with sequences that are similar or identical to endogenous target nucleotide sequences, which facilitates integration of the recombinant nucleic acid at the particular site(s) in the genome containing the endogenous target nucleotide sequences. Nucleic acid sequences adapted for integration via homologous recombination also can include a recognition site for a sequence-specific nuclease. Alternatively, the recognition site for a sequence-specific nuclease can be located in the genome of the cell to be transformed. Donor nucleic acid sequences as described below typically are adapted for integration via homologous recombination.

[0109] In some embodiments, a nucleic acid encoding a selectable marker also can be adapted to integrate via homologous recombination, and thus can be flanked on both sides with sequences that are similar or identical to endogenous sequences within the plant genome (e.g., endogenous sequences at the site of cleavage for a sequence-specific nuclease). In some cases, nucleic acid containing coding sequence for a selectable marker also can include a recognition site for a sequence-specific nuclease. In these embodiments, the recognition site for the sequence-specific nuclease can be the same as or different from that contained within the donor nucleic acid sequence (i.e., can be recognized by the same nuclease as the donor nucleic acid sequence, or recognized by a different nuclease than the donor nucleic acid sequence).

[0110] In some cases, a recombinant nucleic acid sequence can be adapted to integrate into the genome of a cell via site-specific recombination. As used herein, “site-specific” recombination refers to recombination that occurs when a nucleic acid sequence is targeted to a particular site(s) within a genome not by homology between sequences in the recombinant nucleic acid and sequences in the genome, but rather by the action of recombinase enzymes that recognize specific nucleic acid sequences and catalyze the reciprocal exchange of DNA strands between these sites. Site-specific recombination thus refers to the enzyme-mediated cleavage and ligation of two defined nucleotide sequences. Any suitable site-specific recombination system can be used, including, for example, the Cre-lox system or the FLP-FRT system. In such embodiments, a nucleic acid encoding a recombinase enzyme may be introduced into a cell in addition to a donor nucleotide sequence and a nuclease-encoding sequence, and in some cases, a selectable marker sequence. See, e.g., U.S. Pat. No. 4,959,317.

Sequence-Specific Endonucleases

[0111] Sequence-specific nucleases and recombinant nucleic acids encoding the sequence-specific endonucleases are provided herein. The sequence-specific endonucleases can include TAL effector DNA binding domains and endonuclease domains. Thus, nucleic acids encoding such sequence-specific endonucleases can include a nucleotide sequence from a sequence-specific TAL effector linked to a nucleotide sequence from a nuclease.

[0112] TAL effectors are proteins of plant pathogenic bacteria that are injected by the pathogen into the plant cell, where they travel to the nucleus and function as transcription factors to turn on specific plant genes. The primary amino acid sequence of a TAL effector dictates the nucleotide

sequence to which it binds. Thus, target sites can be predicted for TAL effectors, and TAL effectors also can be engineered and generated for the purpose of binding to particular nucleotide sequences, as described herein.

[0113] Fused to the TAL effector-encoding nucleic acid sequences are sequences encoding a nuclease or a portion of a nuclease, typically a nonspecific cleavage domain from a type II restriction endonuclease such as FokI (Kim et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:1156-1160). Other useful endonucleases may include, for example, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, and AlwI. The fact that some endonucleases (e.g., FokI) only function as dimers can be capitalized upon to enhance the target specificity of the TAL effector. For example, in some cases each FokI monomer can be fused to a TAL effector sequence that recognizes a different DNA target sequence, and only when the two recognition sites are in close proximity do the inactive monomers come together to create a functional enzyme. By requiring DNA binding to activate the nuclease, a highly site-specific restriction enzyme can be created.

[0114] A sequence-specific TALEN as provided herein can recognize a particular sequence within a preselected target nucleotide sequence present in a cell. Thus, in some embodiments, a target nucleotide sequence can be scanned for nuclease recognition sites, and a particular nuclease can be selected based on the target sequence. In other cases, a TALEN can be engineered to target a particular cellular sequence. A nucleotide sequence encoding the desired TALEN can be inserted into any suitable expression vector, and can be linked to one or more expression control sequences. For example, a nuclease coding sequence can be operably linked to a promoter sequence that will lead to constitutive expression of the endonuclease in the species of plant to be transformed. Alternatively, an endonuclease coding sequence can be operably linked to a promoter sequence that will lead to conditional expression (e.g., expression under certain nutritional conditions). For example, a cauliflower mosaic virus 35S promoter can be used for constitutive expression. Other constitutive promoters include, without limitation, the nopaline synthase promoter, the ubiquitin promoter, and the actin promoter. In some embodiments, an artificial estrogen-induced promoter for can be used conditional expression, and high levels of transcription can be achieved when a plant is exposed to estrogen. Other conditional promoters that can be used include, for example, heat-inducible heat shock gene promoters, and light-regulated promoters such as that from the gene encoding the large subunit of ribulose biphosphate carboxylase.

[0115] For purposes of therapy, the TAL effector-DNA modifying enzyme of the present document and a pharmaceutically acceptable excipient are administered in a therapeutically effective amount. Such a combination is said to be administered in a “therapeutically effective amount” if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of the recipient. In the present context, an agent is physiologically significant if its presence results in a decrease in the severity of one or more symptoms of the targeted disease and in a genome correction of the lesion or abnormality. Vectors comprising targeting DNA and/or nucleic acid encoding a TAL effector-DNA modifying enzyme can be introduced into a cell by a variety of methods (e.g., injection, direct uptake, projectile bombardment, liposomes, electroporation). TAL effector-DNA modifying

enzymes can be stably or transiently expressed into cells using expression vectors. Techniques of expression in eukaryotic cells are well known to those in the art. (See *Current Protocols in Human Genetics*: Chapter 12 “Vectors For Gene Therapy” and Chapter 13 “Delivery Systems for Gene Therapy”).

[0116] In one further aspect of the present document, the TAL effector-DNA modifying enzyme is substantially non-immunogenic, i.e., engender little or no adverse immunological response. A variety of methods for ameliorating or eliminating deleterious immunological reactions of this sort can be used. In a preferred embodiment, the TAL effector-DNA modifying enzyme is substantially free of N-formyl methionine. Another way to avoid unwanted immunological reactions is to conjugate TAL effector-DNA modifying enzyme to polyethylene glycol (“PEG”) or polypropylene glycol (“PPG”) (preferably of 500 to 20,000 daltons average molecular weight (MW)). Conjugation with PEG or PPG, as described by Davis et al. (U.S. Pat. No. 4,179,337) for example, can provide non-immunogenic, physiologically active, water soluble TAL effector-DNA modifying enzyme conjugates with anti-viral activity. Similar methods also using a polyethylene—polypropylene glycol copolymer are described in Saifer et al. (U.S. Pat. No. 5,006,333).

Donor Vectors

[0117] Also provided herein are recombinant nucleic acids including donor nucleotide sequences. A donor nucleotide sequence can include a variant sequence having one or more modifications (i.e., substitutions, deletions, or insertions) with respect to a preselected target nucleotide sequence found endogenously within the genome of a cell to be transformed (also referred to herein as a “modified target nucleotide sequence”). The variant sequence within the donor nucleic acid typically is flanked on both sides with sequences that are similar or identical to the endogenous target nucleotide sequence within the cell. The flanking sequences can have any suitable length, and typically are at least 50 nucleotides in length (e.g., at least 50 nucleotides, at least 75 nucleotides, at least 100 nucleotides, at least 200 nucleotides, at least 250 nucleotides, at least 300 nucleotides, at least 500 nucleotides, at least 750 nucleotides, at least 1000 nucleotides, from about 50 to about 5000 nucleotides, from about 100 to 2500 nucleotides, from about 100 to about 1000 nucleotides, from about 100 to 500 nucleotides, from about 200 to about 500 nucleotides, or from about 250 to 400 nucleotides). Thus, homologous recombination can occur between the recombinant donor nucleic acid construct and the endogenous target on both sides of the variant sequence, such that the resulting cell’s genome contains the variant sequence within the context of endogenous sequences from, for example, the same gene. A donor nucleotide sequence can be generated to target any suitable sequence within a genome. In a plant, for example, a donor nucleotide sequence can be targeted to a lipid biosynthetic gene, carbohydrate biosynthetic gene, seed storage protein gene, disease or pest resistance gene, stress tolerance gene, drought tolerance gene, or a gene that produces an anti-nutritional. In addition, the donor nucleotide sequence contains a recognition site for a sequence-specific nuclease, as described herein.

Selectable Markers

[0118] Some of the methods provided herein include the use of a third recombinant nucleic acid encoding a selectable

or screenable marker. A nucleotide sequence encoding a polypeptide that results in a selectable trait can be incorporated into an expression vector containing one or more expression control sequences. For example, an expression vector can include sequence encoding a selectable marker operably linked to a promoter sequence that will lead to constitutive expression in the plant cell to be transformed. Suitable selectable markers can include, without limitation, polypeptides conferring resistance to an antibiotic such as kanamycin, G418, bleomycin, ampicillin, or hygromycin, or an herbicide such as glufosinate, chlorosulfuron, or phosphinothricin.

[0119] In embodiments for use in plants, for example, a selectable marker can confer resistance to an herbicide that inhibits the growing point or meristem, such as an imidazolinone or a sulfonamide. Exemplary polypeptides in this category code for mutant ALS and AHAS enzymes as described, for example, in U.S. Pat. Nos. 5,767,366 and 5,928,937. U.S. Pat. Nos. 4,761,373 and 5,013,659 are directed to plants resistant to various imidazolinone or sulfonamide herbicides. U.S. Pat. No. 4,975,374 relates to plant cells and plants containing a gene encoding a mutant glutamine synthetase (GS) resistant to inhibition by herbicides that are known to inhibit GS, e.g., phosphinothricin and methionine sulfoximine. U.S. Pat. No. 5,162,602 discloses plants resistant to inhibition by cyclohexanedione and aryloxyphenoxypropanoic acid herbicides. The resistance is conferred by an altered acetyl coenzyme A carboxylase (ACCase).

[0120] Polypeptides for resistance to glyphosate (sold under the trade name Roundup®) also are suitable for use in plants. See, for example, U.S. Pat. Nos. 4,940,835 and 4,769,061. U.S. Pat. No. 5,554,798 discloses transgenic glyphosate resistant maize plants, in which resistance is conferred by an altered 5-enolpyruvyl-3-phosphoshikimate (EPSP) synthase. Such polypeptides can confer resistance to glyphosate herbicidal compositions including, without limitation, glyphosate salts such as the trimethylsulphonium salt, the isopropylamine salt, the sodium salt, the potassium salt and the ammonium salt. See, e.g., U.S. Pat. Nos. 6,451,735 and 6,451,732.

[0121] Polypeptides for resistance to phosphonate compounds such as glufosinate ammonium or phosphinothricin, and pyridinoxy or phenoxy propionic acids and cyclohexones also are suitable. See, for example, European Publication No. 0 242 246, as well as U.S. Pat. Nos. 5,879,903, 5,276,268, and 5,561,236.

[0122] Other herbicides include those that inhibit photosynthesis, such as triazine and benzonitrile (nitrilase). See, e.g., U.S. Pat. No. 4,810,648. Other herbicides include 2,2-dichloropropionic acid, sethoxydim, haloxyfop, imidazolinone herbicides, sulfonamide herbicides, triazolopyrimidine herbicides, s-triazine herbicides and bromoxynil. Also suitable are herbicides that confer resistance to a protox enzyme. See, e.g., U.S. Patent Publication No. 20010016956 and U.S. Pat. No. 6,084,155.

[0123] In some embodiments, a recombinant nucleic acid encoding a selectable marker can be adapted to integrate into the genome of a cell (e.g., a plant cell or an animal cell) by site-specific recombination. For example, a sequence encoding a selectable marker can be flanked by recognition sequences for a recombinase such as, e.g., Cre or FLP. In other embodiments, a recombinant nucleic acid encoding a selectable marker can be adapted for integration into a plant genome by homologous recombination. In such nucleic acids, the sequence encoding the selectable marker can be

flanked by sequences that are similar or identical to endogenous nucleotide sequences found within the genome of the plant cell into which the recombinant nucleic acid is to be introduced. At least one of the endogenous sequences can be at the cleavage site for a sequence-specific nuclease. The nucleic acid encoding the selectable marker also can contain a recognition site for a sequence-specific nuclease. The nuclease can be the same sequence-specific nuclease as that which is targeted to the donor nucleotide sequence, or a sequence-specific nuclease that is different from that targeted to the donor nucleotide sequence. In still other embodiments, a recombinant nucleic acid encoding a selectable marker can be adapted for integration into the genome of a plant cell by illegitimate recombination. Such nucleic acids typically lack the flanking sequences and nuclease recognition sites that are contained within nucleic acids adapted for homologous or site-specific recombination as described herein.

Methods

[0124] One or more of the constructs provided herein can be used to transform cells and/or a DNA modifying enzyme can be introduced into cells, such that a genetically modified organism (e.g., a plant or an animal) is generated. Thus, genetically modified organisms and cells containing the nucleic acids and/or polypeptides described herein also are provided. In some embodiments, a transformed cell has a recombinant nucleic acid construct integrated into its genome, i.e., can be stably transformed. Stably transformed cells typically retain the introduced nucleic acid sequence with each cell division. A construct can integrate in a homologous manner, such that a nucleotide sequence endogenous to the transformed cell is replaced by the construct, where the construct contains a sequence that corresponds to the endogenous sequence, but that contains one or more modifications with respect to the endogenous sequence. It is noted that while a plant or animal containing such a modified endogenous sequence may be termed a “genetically modified organism” (GMO) herein, the modified endogenous sequence is not considered a transgene. A construct also can integrate in an illegitimate manner, such that it integrates randomly into the genome of the transformed cell.

[0125] Alternatively, a cell can be transiently transformed, such that the construct is not integrated into its genome. For example, a plasmid vector containing a TALEN coding sequence can be introduced into a cell, such that the TALEN coding sequence is expressed but the vector is not stably integrated in the genome. Transiently transformed cells typically lose some or all of the introduced nucleic acid construct with each cell division, such that the introduced nucleic acid cannot be detected in daughter cells after sufficient number of cell divisions. Nevertheless, expression of the TALEN coding sequence is sufficient to achieve homologous recombination between a donor sequence and an endogenous target sequence. Both transiently transformed and stably transformed cells can be useful in the methods described herein.

[0126] With particular respect to genetically modified plant cells, cells used in the methods described herein can constitute part or all of a whole plant. Such plants can be grown in a manner suitable for the species under consideration, either in a growth chamber, a greenhouse, or in a field. Genetically modified plants can be bred as desired for a particular purpose, e.g., to introduce a recombinant nucleic acid into other lines, to transfer a recombinant nucleic acid to other species or for further selection of other desirable traits. Alternatively,

genetically modified plants can be propagated vegetatively for those species amenable to such techniques. Progeny includes descendants of a particular plant or plant line. Progeny of an instant plant include seeds formed on F_1 , F_2 , F_3 , F_4 , F_5 , F_6 and subsequent generation plants, or seeds formed on BC_1 , BC_2 , BC_3 , and subsequent generation plants, or seeds formed on F_1BC_1 , F_1BC_2 , F_1BC_3 , and subsequent generation plants. Seeds produced by a genetically modified plant can be grown and then selfed (or outcrossed and selfed) to obtain seeds homozygous for the nucleic acid construct.

[0127] Genetically modified cells (e.g., plant cells or animal cells) can be grown in suspension culture, or tissue or organ culture, if desired. For the purposes of the methods provided herein, solid and/or liquid tissue culture techniques can be used. When using solid medium, cells can be placed directly onto the medium or can be placed onto a filter film that is then placed in contact with the medium. When using liquid medium, cells can be placed onto a floatation device, e.g., a porous membrane that contacts the liquid medium. Solid medium typically is made from liquid medium by adding agar. For example, a solid medium can be Murashige and Skoog (MS) medium containing agar and a suitable concentration of an auxin, e.g., 2,4-dichlorophenoxyacetic acid (2,4-D), and a suitable concentration of a cytokinin, e.g., kinetin.

[0128] A cell can be transformed with one recombinant nucleic acid construct or with a plurality (e.g., 2, 3, 4, or 5) of recombinant nucleic acid constructs. If multiple constructs are utilized, they can be transformed simultaneously or sequentially. Techniques for transforming a wide variety of species are known in the art. The polynucleotides and/or recombinant vectors described herein can be introduced into the genome of a host using any of a number of known methods, including electroporation, microinjection, and biolistic methods. Alternatively, polynucleotides or vectors can be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. Such *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well known in the art. Other gene transfer and transformation techniques include protoplast transformation through calcium or PEG, electroporation-mediated uptake of naked DNA, liposome-mediated transfection, electroporation, viral vector-mediated transformation, and microprojectile bombardment (see, e.g., U.S. Pat. Nos. 5,538,880, 5,204,253, 5,591,616, and 6,329,571). If a plant cell or tissue culture is used as the recipient tissue for transformation, plants can be regenerated from transformed cultures using techniques known to those skilled in the art.

[0129] In some embodiments, a DNA modifying enzyme (e.g., a TALEN) can be directly introduced into a cell. For example, a polypeptide can be introduced into a cell by mechanical injection, by delivery via a bacterial type III secretion system, by electroporation, or by *Agrobacterium* mediated transfer. See, e.g., Vergunst et al. (2000) *Science* 290:979-982 for a discussion of the *Agrobacterium* VirB/D4 transport system, and its use to mediate transfer of a nucleoprotein T complex into plant cells.

[0130] With further respect to plants, the polynucleotides, vectors and polypeptides described herein can be introduced into a number of monocotyledonous and dicotyledonous plants and plant cell systems, including dicots such as safflower, alfalfa, soybean, coffee, amaranth, rapeseed (high erucic acid and canola), peanut or sunflower, as well as monocots such as oil palm, sugarcane, banana, sudangrass, corn,

wheat, rye, barley, oat, rice, millet, or sorghum. Also suitable are gymnosperms such as fir and pine.

[0131] Thus, the methods described herein can be utilized with dicotyledonous plants belonging, for example, to the orders *Magniolales*, *Illiciales*, *Laurales*, *Piperiales*, *Aristolochiales*, *Nymphaeales*, *Ranunculales*, *Papeverales*, *Sarracenaceae*, *Trochodendrales*, *Hamamelidales*, *Eucomiales*, *Leitneriales*, *Myricales*, *Fagales*, *Casuarinales*, *Caryophyllales*, *Batales*, *Polygonales*, *Plumbaginales*, *Dilleniales*, *Theales*, *Malvales*, *Urticales*, *Lecythidales*, *Violales*, *Salicales*, *Capparales*, *Ericales*, *Diapensales*, *Ebenales*, *Primulales*, *Rosales*, *Fabales*, *Podostemales*, *Haloragales*, *Myrtales*, *Cornales*, *Proteales*, *Santales*, *Rafflesiales*, *Celastrales*, *Euphorbiales*, *Rhamnales*, *Sapindales*, *Juglandales*, *Geraniales*, *Polygalales*, *Umbellales*, *Gentianales*, *Polemoniales*, *Lamiales*, *Plantaginales*, *Scrophulariales*, *Campanulales*, *Rubiales*, *Dipsacales*, and *Asterales*. The methods described herein also can be utilized with monocotyledonous plants such as those belonging to the orders *Alismatales*, *Hydrocharitales*, *Najadales*, *Triuridales*, *Commelinales*, *Eriocaulales*, *Restionales*, *Poales*, *Juncales*, *Cyperales*, *Typhales*, *Bromeliales*, *Zingiberales*, *Areciales*, *Cyclanthales*, *Pandanales*, *Arales*, *Lilliales*, and *Orchidales*, or with plants belonging to *Gymnospermae*, e.g., *Pinales*, *Ginkgoales*, *Cycadales* and *Gnetales*.

[0132] The methods can be used over a broad range of plant species, including species from the dicot genera *Atropa*, *Alseodaphne*, *Anacardium*, *Arachis*, *Beilschmiedia*, *Brassica*, *Carthamus*, *Cocculus*, *Croton*, *Cucumis*, *Citrus*, *Citrullus*, *Capsicum*, *Catharanthus*, *Cocos*, *Coffea*, *Cucurbita*, *Daucus*, *Duguetia*, *Eschscholzia*, *Ficus*, *Fragaria*, *Glaucium*, *Glycine*, *Gossypium*, *Helianthus*, *Hevea*, *Hyoscyamus*, *Lactuca*, *Landolphia*, *Linum*, *Litsea*, *Lycopersicon*, *Lupinus*, *Manihot*, *Majorana*, *Malus*, *Medicago*, *Nicotiana*, *Olea*, *Parthenium*, *Papaver*, *Persea*, *Phaseolus*, *Pistacia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Senecio*, *Sinomenium*, *Stephania*, *Sinapis*, *Solanum*, *Theobroma*, *Trifolium*, *Trigonella*, *Vicia*, *Vinca*, *Vitis*, and *Vigna*; the monocot genera *Allium*, *Andropogon*, *Aragrostis*, *Asparagus*, *Avena*, *Cynodon*, *Elaeis*, *Festuca*, *Festulolium*, *Heterocallis*, *Hordeum*, *Lemna*, *Lolium*, *Musa*, *Oryza*, *Panicum*, *Pannisetum*, *Phleum*, *Poa*, *Secale*, *Sorghum*, *Triticum*, and *Zea*; or the gymnosperm genera *Abies*, *Cunninghamia*, *Picea*, *Pinus*, and *Pseudotsuga*.

[0133] A transformed cell, callus, tissue, or plant can be identified and isolated by selecting or screening the engineered cells for particular traits or activities, e.g., those encoded by marker genes or antibiotic resistance genes. Such screening and selection methodologies are well known to those having ordinary skill in the art. In addition, physical and biochemical methods can be used to identify transformants. These include Southern analysis or PCR amplification for detection of a polynucleotide; Northern blots, S1 RNase protection, primer-extension, or RT-PCR amplification for detecting RNA transcripts; enzymatic assays for detecting enzyme or ribozyme activity of polypeptides and polynucleotides; and protein gel electrophoresis, Western blots, immunoprecipitation, and enzyme-linked immunoassays to detect polypeptides. Other techniques such as in situ hybridization, enzyme staining, and immunostaining also can be used to detect the presence or expression of polypeptides and/or polynucleotides. Methods for performing all of the referenced techniques are well known. Polynucleotides that are

stably incorporated into plant cells can be introduced into other plants using, for example, standard breeding techniques.

[0134] In the context of the present document, “eukaryotic cells” refer to a fungal, yeast, plant or animal cell or a cell line derived from the organisms listed below and established for in vitro culture. More preferably, the fungus can be of the genus *Aspergillus*, *Penicillium*, *Acremonium*, *Trichoderma*, *Chrysosporium*, *Mortierella*, *Cluyveromyces* or *Pichia*. More preferably, the fungus can be of the species *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus oryzae*, *Aspergillus terreus*, *Penicillium chrysogenum*, *Penicillium citrinum*, *Acremonium Chrysogenum*, *Trichoderma reesei*, *Mortierella alpine*, *Chrysosporium lucknowense*, *Cluyveromyces lactis*, *Pichia pastoris* or *Pichia ciferrii*.

[0135] The plant can be of the genus *Arabidopsis*, *Nicotiana*, *Solanum*, *Lactuca*, *Brassica*, *Oryza*, *Asparagus*, *Pisum*, *Medicago*, *Zea*, *Hordeum*, *Secale*, *Triticum*, *Capsicum*, *Cucumis*, *Cucurbita*, *Citrullis*, *Citrus*, or *Sorghum*. More preferably, the plant can be of the species *Arabidopsis thaliana*, *Nicotiana tabaccum*, *Solanum lycopersicum*, *Solanum tuberosum*, *Solanum melongena*, *Solanum esculentum*, *Lactuca saliva*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Oryza glaberrima*, *Oryza sativa*, *Asparagus officinalis*, *Pisum sativum*, *Medicago sativa*, *Zea mays*, *Hordeum vulgare*, *Secale cereal*, *Triticum aestivum*, *Triticum durum*, *Capsicum sativum*, *Cucurbita pepo*, *Citrullus lanatus*, *Cucumis melo*, *Citrus aurantifolia*, *Citrus maxima*, *Citrus medica*, or *Citrus reticulata*.

[0136] The animal cell can be of the genus *Homo*, *Rattus*, *Mus*, *Sus*, *Bos*, *Danio*, *Canis*, *Felis*, *Equus*, *Salmo*, *Oncorhynchus*, *Gallus*, *Meleagris*, *Drosophila*, or *Caenorhabditis*; more preferably, the animal cell can be of the species *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Sus scrofa*, *Bos taurus*, *Danio rerio*, *Canis lupus*, *Felis catus*, *Equus caballus*, *Oncorhynchus mykiss*, *Gallus gallus*, or *Meleagris gallopavo*; the animal cell can be a fish cell from *Salmo salar*, Teleost fish or zebrafish species as non-limiting examples. The animal cell also can be an insect cell from *Drosophila melanogaster* as a non-limiting example; the animal cell can also be a worm cell from *Caenorhabditis elegans* as a non-limiting example.

[0137] In the present document, the cell can be a plant cell, a mammalian cell, a fish cell, an insect cell or cell lines derived from these organisms for in vitro cultures or primary cells taken directly from living tissue and established for in vitro culture. As non limiting examples cell lines can be selected from the group consisting of CHO-K1 cells; HEK293 cells; Caco2 cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; CHO-S cells; DG44 cells; K-562 cells; U-937 cells; MRCS cells; IMR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HT-1080 cells; HCT-116 cells; Hu-h7 cells; Huvec cells; Molt 4 cells.

[0138] All these cell lines can be modified by the method of the present document to provide cell line models to produce, express, quantify, detect, study a gene or a protein of interest; these models can also be used to screen biologically active molecules of interest in research and production in various fields such as chemical, biofuels, therapeutics and agronomy as non-limiting examples.

[0139] The present document also provides methods for harnessing the sequence-specific DNA binding domains within TAL effectors to, for example, alter the genetic material within cells, to modulate gene expression, and to target

pathogenic sequences in, e.g., anti-viral therapies. For example, in some embodiments, the present document provides methods for modifying cellular genetic material. In some embodiments, the methods include introducing a polypeptide containing a TAL effector DNA binding domain, or a nucleic acid encoding such a polypeptide, into a cell. The TAL effector DNA binding domain can be fused to all or a portion of a DNA modifying enzyme (e.g., an endonuclease). In some embodiments, the methods include introducing two or more recombinant nucleic acids into a cell. A first recombinant nucleic acid contains a donor nucleotide sequence that includes one or more modifications (i.e., substitutions, deletions, or insertions) with respect to a corresponding, pre-selected target nucleotide sequence found in the cell. The donor nucleotide sequence can undergo homologous recombination with the endogenous target nucleotide sequence, such that the endogenous sequence or a portion thereof is replaced with the donor sequence or a portion thereof. The target nucleotide sequence typically includes a recognition site for a sequence-specific TALEN. In some cases, a target nucleotide sequence can include recognition sites for two or more distinct TALENs (e.g., two opposed target sequences that are distinct, such that TALENs having distinct DNA sequence binding specificity can be used). In such cases, the specificity of DNA cleavage can be increased as compared to cases in which only one target sequence (or multiple copies of the same target sequence) is used.

[0140] A second recombinant nucleic acid contains a nucleotide sequence encoding a sequence specific TALEN that binds to the recognition site in the target nucleotide sequence. In some cases, the donor nucleotide sequence and the nucleotide sequence encoding the sequence-specific nuclease can be contained in the same nucleic acid construct. Alternatively, the donor nucleotide sequence and the TALEN coding sequence can be contained in separate constructs, or the TALEN polypeptide can be produced and introduced directly into a cell.

[0141] In some embodiments, a third recombinant nucleic acid containing a nucleotide sequence encoding a selectable marker also may be used. The second and third recombinant nucleic acids may undergo recombination with endogenous sequences and thus integrate into the genome of the cell. These recombination events can be illegitimate (i.e., random), or they can occur through homologous recombination or through site-specific recombination. The recombinant nucleic acids can be simultaneously or sequentially transformed into the cell, and can be linearized prior to transformation.

[0142] When the cell is a plant cell, the methods provided herein can further include steps such as generating a plant containing the transformed cell, generating progeny of the plant, selecting or screening for plants expressing the selectable marker (if included), generating progeny of the selected plants, and testing the plants (e.g., tissue, seed, precursor cells, or whole plants) or progeny of the plants for recombination at the target nucleotide sequence. In some cases, the methods can include out-crossing the selected plants to remove the selectable marker, and/or screening the selected or out-crossed plants for the absence of the sequence-specific nuclease.

[0143] In some embodiments, the present document provides methods for modifying the genetic material of a cell, e.g., a prokaryotic cell, an animal cell, or a plant cell. The methods can include introducing into the cell a first recom-

binant nucleic acid containing a modified target nucleotide sequence that includes one or more modifications in nucleotide sequence with respect to a corresponding target nucleotide sequence present in the cell, as well as a recognition site for a sequence-specific TALEN, and a second recombinant nucleic acid containing a nucleotide sequence encoding the sequence-specific TALEN. When the cell is a plant cell, a plant containing the cell can be generated, and cells, seed, or tissue obtained from the plant (or progeny thereof) can be analyzed for recombination at the target nucleotide sequence. The first and second recombinant nucleic acids can be simultaneously or serially transformed into the cell, and one or both may be linearized prior to transformation. In some cases, the first and second recombinant nucleic acids can be present in the same construct.

[0144] In some cases, the method also can include introducing into the cell a third recombinant nucleic acid containing a nucleotide sequence encoding a selectable marker, and determining whether the cell, an organism generated from the cell, or progeny thereof expresses the selectable marker. The method further can include screening the cell, the organism or progeny thereof for the absence of the selectable marker. The nucleotide sequence encoding the selectable marker may or may not be flanked on both sides by nucleotide sequences that are similar or identical to nucleotide sequences endogenous to the cell at the site of cleavage for a second sequence-specific nuclease, or by recognition sites for a sequence-specific recombinase. In some cases, the method also can include the step of out-crossing the organism. Progeny of the out-cross can be screened for the absence of the selectable marker.

[0145] The present document also provides methods for modifying the genetic material of a cell (e.g., a plant cell or an animal cell), comprising providing a cell containing a target DNA sequence, e.g., a chromosomal, mitochondrial, or chloroplast sequence, in which it is desired to have homologous recombination occur, providing a TALEN that contains a DNA modifying enzyme domain (e.g., an endonuclease domain) and a TAL effector domain having a plurality of TAL effector repeats that, in combination, bind to a specific nucleotide sequence within the target DNA sequence, providing a nucleic acid containing a sequence homologous to at least a portion of the target DNA, and contacting the target DNA sequence in the cell with the TAL endonuclease such that both strands of a nucleotide sequence within or adjacent to the target DNA sequence in the cell are cleaved. Such cleavage can enhance the frequency of homologous recombination at the target DNA sequence. The target DNA sequence can be endogenous to the cell. The methods can include introducing into the cell a vector containing a cDNA encoding the TAL endonuclease, and expressing a TAL endonuclease protein in the cell. In some cases, the TAL endonuclease protein itself can be introduced into the cell, for example, by mechanical injection, by delivery via a bacterial type III secretion system, by electroporation, or by *Agrobacterium* mediated transfer.

[0146] The methods described herein can be used in a variety of situations. In agriculture, for example, methods described herein are useful to facilitate homologous recombination at a target site can be used to remove a previously integrated transgene (e.g., a herbicide resistance transgene) from a plant line, variety, or hybrid. The methods described herein also can be used to modify an endogenous gene such that the enzyme encoded by the gene confers herbicide resistance, e.g., modification of an endogenous 5-enolpyruvyl

shikimate-3-phosphate (EPSP) synthase gene such that the modified enzyme confers resistance to glyphosate herbicides. As another example, the methods described herein are useful to facilitate homologous recombination at regulatory regions for one or more endogenous genes in a plant or mammal metabolic pathway (e.g., fatty acid biosynthesis), such that expression of such genes is modified in a desired manner. The methods described herein are useful to facilitate homologous recombination in an animal (e.g., a rat or a mouse) in one or more endogenous genes of interest involved in, as non-limiting examples, metabolic and internal signaling pathways such as those encoding cell-surface markers, genes identified as being linked to a particular disease, and any genes known to be responsible for a particular phenotype of an animal cell.

[0147] The present document also provides methods for designing sequence-specific TAL effectors capable of interacting with particular DNA sequences (e.g., TALENs capable of cleaving DNA at specific locations). The methods can include identifying a target nucleotide sequence (e.g., an endogenous chromosomal sequence, a mitochondrial DNA sequence, or a chloroplast DNA sequence) at which it is desired to have TAL effector binding (e.g., a sequence adjacent to a second nucleotide sequence at which it is desired to introduce a double-stranded cut), and designing a sequence specific TAL effector that contains a plurality of DNA binding repeats that, in combination, bind to the target sequence. As described herein, TAL effectors include a number of imperfect repeats that determine the specificity with which they interact with DNA. Each repeat binds to a single base, depending on the particular di-amino acid sequence at residues 12 and 13 of the repeat. Thus, by engineering the repeats within a TAL effector (e.g., using standard techniques or the techniques described herein), particular DNA sites can be targeted. Such engineered TAL effectors can be used, for example, as transcription factors targeted to particular DNA sequences. A diagram of a generic TAL effector is shown in FIG. 1A, with the repeat region indicated by open boxes, and the RVD in the representative repeat sequence (SEQ ID NO:1) underlined.

[0148] Examples of RVDs and their corresponding target nucleotides are shown in Table 1A (See, also, PCT Publication No. WO2010/079430).

TABLE 1A

RVD	Nucleotide
HD	C
NG	T
NI	A
NN	G or A
NS	A or C or G
N*	C or T
HG	T
H*	T
IG	T

*Denotes a gap in the repeat sequence corresponding to a lack of an amino acid residue at the second position of the RVD.

Other RVDs and their corresponding target nucleotides are shown in Table 1B.

TABLE 1B

RVD	Nucleotide
HA	C
ND	C
NK	G

TABLE 1B-continued

RVD	Nucleotide
HI	C
HN	G
NA	G
SN	G or A
YG	T

[0149] When it is desired to have sequence-specific DNA cleavage, for example, a sequence-specific TALEN can be designed to contain (a) a plurality of DNA binding repeat domains that, in combination, bind to the endogenous chromosomal nucleotide sequence, and (b) an endonuclease that generates a double-stranded cut at the second nucleotide sequence. Such sequence-specific DNA cleavage can be useful to enhance homologous recombination, as described herein. Other uses for TALENs include, for example, as therapeutics against viruses. TALENs can be engineered to target particular viral sequences, cleaving the viral DNA and reducing or abolishing virulence.

[0150] The materials and methods provided herein can be used to modify the sequence of a particular gene in a targeted manner. A gene may contain a plurality of sequences to which an engineered TAL effector could be targeted. As described herein, however, certain target sequences may be more effectively targeted. For example, as set forth in Example 9, sequences having particular characteristics may be more effectively targeted by TAL effectors. Thus, the methods provided herein can include identifying target sequences that meet particular criteria. These include sequences that: i) have a minimum length of 15 bases and an orientation from 5' to 3' with a T immediately preceding the site at the 5' end; ii) do not have a T in the first (5') position or an A in the second position; iii) end in T at the last (3') position and do not have a G at the next to last position; and iv) have a base composition of 0-63% A, 11-63% C, 0-25% G, and 2-42% T.

[0151] Since TALENs as described herein generally work as dimers, some embodiments of the methods provided herein can include identifying a first genomic nucleotide sequence and a second genomic nucleotide sequence in a cell, wherein the first and second nucleotide sequences meet at least one of the criteria set forth above and are separated by 15-18 bp. In some cases, one TALEN polypeptide can bind to each nucleotide sequences, and the endonuclease contained in the TALEN can cleave within the 15-18 bp spacer.

[0152] The present document also provides methods for generating genetically modified animals into which a desired nucleic acid has been introduced. Such methods can include obtaining a cell containing an endogenous chromosomal target DNA sequence into which it is desired to introduce the nucleic acid, introducing into the cell a TALEN to generate a double-stranded cut within the endogenous chromosomal target DNA sequence, introducing into the cell an exogenous nucleic acid containing a sequence homologous to at least a portion of the endogenous chromosomal target DNA, where the introduction is done under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the endogenous chromosomal target DNA, and generating an animal from the primary cell in which homologous recombination has occurred. The homologous nucleic acid can include, e.g., a nucleotide sequence that disrupts a gene after homologous recombination, a nucleotide sequence that replaces a gene after homologous recombination, a nucle-

otide sequence that introduces a point mutation into a gene after homologous recombination, or a nucleotide sequence that introduces a regulatory site after homologous recombination.

[0153] The methods provided herein also can be used to generate genetically modified plants in which a desired nucleic acid has been introduced. Such methods can include obtaining a plant cell containing an endogenous target DNA sequence into which it is desired to introduce the nucleic acid, introducing a TALEN to generate a double-stranded cut within the endogenous target DNA sequence, introducing into the plant cell an exogenous nucleic acid containing a sequence homologous to at least a portion of the endogenous target DNA, where the introducing is under conditions that permit homologous recombination to occur between the exogenous nucleic acid and the endogenous target DNA, and generating a plant from the plant cell in which homologous recombination has occurred.

[0154] The DNA in cells generated by the TALEN-facilitated homologous recombination methods provided herein is modified, as compared to cells that have not undergone such methods, and cells containing the modified DNA are referred to as “genetically modified.” It is noted, however, that organisms containing such cells may not be considered GMO for regulatory purposes, since such a modification involves a homologous recombination and not random integration of a transgene. Thus, using the TALEN-facilitated methods described herein to generate genetic modifications may be advantageous in that, for example, standard regulatory procedures along with their associated time and cost may be avoided.

[0155] Other methods of targeted genetic recombination, as provided herein, can include introducing into a cell (e.g., a plant cell, insect cell, teleost fish cell, or animal cell) a nucleic acid molecule encoding a TALEN targeted to a selected DNA target sequence, inducing expression of the TALEN within the cell, and identifying a recombinant cell in which the selected DNA target sequence exhibits a mutation (e.g., a deletion of genetic material, an insertion of genetic material, or both a deletion and an insertion of genetic material). A donor DNA also can be introduced into the cell.

[0156] In some embodiments, a monomeric TALEN can be used. TALENs as described herein typically function as dimers across a bipartite recognition site with a spacer, such that two TAL effector domains are each fused to a catalytic domain of the FokI restriction enzyme, the DNA recognition sites for each resulting TALEN are separated by a spacer sequence, and binding of each TALEN monomer to the recognition site allows FokI to dimerize and create a double-strand break within the spacer (see, e.g., Moscou and Bogdanove (2009) *Science* 326:1501). Monomeric TALENs also can be constructed, however, such that single TAL effectors are fused to a nuclease that does not require dimerization to function. One such nuclease, for example, is a single-chain variant of FokI in which the two monomers are expressed as a single polypeptide (Minczuk et al. (2008) *Nucleic Acids Res.* 36:3926-3938). Other naturally occurring or engineered monomeric nucleases also can serve this role. The DNA recognition domain used for a monomeric TALEN can be derived from a naturally occurring TAL effector. Alternatively, the DNA recognition domain can be engineered to recognize a specific DNA target. Engineered single-chain TALENs may be easier to construct and deploy, as they require only one engineered DNA recognition domain.

[0157] In some embodiments, a dimeric DNA sequence-specific nuclease can be generated using two different DNA binding domains (e.g., one TAL effector binding domain and one binding domain from another type of molecule). As set forth above, the TALENs described herein typically function as dimers across a bipartite recognition site with a spacer. This nuclease architecture also can be used for target-specific nucleases generated from, for example, one TALEN monomer and one zinc finger nuclease monomer. In such cases, the DNA recognition sites for the TALEN and zinc finger nuclease monomers can be separated by a spacer of appropriate length. Binding of the two monomers can allow FokI to dimerize and create a double-strand break within the spacer sequence. DNA binding domains other than zinc fingers, such as homeodomains, myb repeats or leucine zippers, also can be fused to FokI and serve as a partner with a TALEN monomer to create a functional nuclease.

[0158] In some embodiments, a TAL effector can be used to target other protein domains (e.g., non-nuclease protein domains) to specific nucleotide sequences. For example, a TAL effector can be linked to a protein domain from, without limitation, a DNA interacting enzyme (e.g., a methylase, a topoisomerase, an integrase, a transposase, or a ligase), a transcription activator or repressor, or a protein that interacts with or modifies other proteins such as histones. Applications of such TAL effector fusions include, for example, creating or modifying epigenetic regulatory elements, making site-specific insertions, deletions, or repairs in DNA, controlling gene expression, and modifying chromatin structure.

[0159] In some embodiments, the spacer of the target sequence can be selected or varied to modulate TALEN specificity and activity. The results presented herein for TALENs that function as dimers across a bipartite recognition site with a spacer demonstrate that TALENs can function over a range of spacer lengths, and that the activity of TALENs varies with spacer length. See, e.g., Example 6 below. The flexibility in spacer length indicates that spacer length can be chosen to target particular sequences (e.g., in a genome) with high specificity. Further, the variation in activity observed for different spacer lengths indicates that spacer length can be chosen to achieve a desired level of TALEN activity.

[0160] In some embodiments, TALEN activity can be modulated by varying the number and composition of repeats within the DNA binding domain(s). As described in Example 7 herein, for example, a PthXoI-based TALEN showed greater activity than an AvrBs3-based TALEN. PthXoI differs from AvrBs3 both in the number and RVD composition of its repeats. In addition, the naturally occurring DNA recognition sites for these proteins differ in their divergence from the respective recognition sequences predicted based on the TAL effector DNA cipher described by Moscou and Bogdanove (supra). Further, several custom TALENs of the same length (12 RVDs) but with differing RVD composition differed in their activity, and a 13 RVD custom TALEN had higher activity than a 12 RVD custom TALEN. Thus, not only can TALENs be engineered to recognize a DNA sequence of interest, but (1) the number of repeats can be varied to modulate activity, (2) different binding sites can be selected to achieve different levels of activity, and (3) the composition of RVDs and their fit to the target site (according to the cipher) can be varied to modulate TALEN activity.

[0161] When the TALEN is in a heterodimeric form, for instance with two different monomers including each a TAL effector domain and a FokI nuclease catalytic domain, the

RVDs can be found in equivalent number in each of the two TAL effector domains, or each domain can display different numbers of RVDs. For instance, if a total of 22 RVDs is used to bind DNA in a particular heterodimeric TALEN, 11 repeats can be found in each of the two TAL effector domains; alternatively, 10 repeats can be found in one of the two TAL effector domains and 12 in the other. The present document also encompasses TALEN with DNA modifying enzyme domain which functions as a monomer. In this case, all the RVDs can be found in a single TAL effector domain, which is fused to the monomeric enzyme. In this case, in order to have efficient binding, the number of RVDs must be equivalent to the total number of RVDs that would be found in an equivalent dimeric TALEN. For example, instead of having 10 repeats on two different TAL effector domains (as in the case for a dimeric TALEN), one would have 20 repeats in a single TAL effector domain (as in the case for a monomeric TALEN).

[0162] In a further aspect, the total number of repeats within the dimeric or monomeric TALEN is at least 14. In another further aspect, the total number of repeats within the dimeric or monomeric TALEN is at least 20. In another further aspect, the total number of repeats within the dimeric or monomeric TALEN is at least 24. In another further aspect, the total number of repeats within the dimeric or monomeric TALEN is at least 30.

[0163] This patent application also provides methods for generating TAL effector proteins having enhanced targeting capacity for a target DNA. The methods can include, for example, generating a nucleic acid encoding a TAL effector that has a DNA binding domain with a plurality of DNA binding repeats, each repeat containing a RVD that determines recognition of a base pair in the target DNA, where each DNA binding repeat is responsible for recognizing one base pair in the target DNA. As described in Example 12 below, relaxing the requirement for T at position -1 of the binding site may enhance the targeting capacity for engineered TAL effector proteins. Thus, generating a TAL effector encoding nucleic acid can include incorporating a nucleic acid encoding a variant 0th DNA binding repeat sequence with specificity for A, C, or G, thus eliminating the requirement for T at position -1 of the binding site.

[0164] In addition, methods are provided herein for generating TAL effectors having enhanced targeting capacity for a target DNA. Such methods can include generating a nucleic acid encoding a TAL effector that comprises DNA binding domain having a plurality of DNA binding repeats, each repeat containing a RVD that determines recognition of a base pair in the target DNA. As described in Example 12 below, the specificity of NN (the most common RVD that recognizes G) appears to be generally weak and can vary with context, but certain RVDs may have enhanced specificity for G. Thus, methods provided herein can include using alternate RVDs that may have more robust specificity for G. For example, one or more RVDs selected from the group consisting of RN, R*, NG, NH, KN, K*, NA, NT, DN, D*, NL, NM, EN, E*, NV, NC, QN, Q*, NR, NP, HN, H*, NK, NY, SN, S*, ND, NR, TN, T*, NE, NF, YN, Y*, and NQ can be used, where the asterisk indicates a gap at the second position of the RVD.

Articles of Manufacture

[0165] The present document also provides articles of manufacture containing, for example, nucleic acid molecules encoding TALENs, TALEN polypeptides, compositions con-

taining such nucleic acid molecules or polypeptides, or TAL endonuclease engineered cell lines. Such items can be used, for example, as research tools, or therapeutically.

[0166] In some embodiments, an article of manufacture can include seeds from plants generated using methods provided herein. The seeds can be conditioned using means known in the art and packaged using packaging material well known in the art to prepare an article of manufacture. A package of seed can have a label e.g., a tag or label secured to the packaging material, a label printed on the packaging material or a label inserted within the package. The label can indicate that the seeds contained within the package can produce a crop of genetically modified plants, and can describe the traits that are altered by the genetic modification, relative to unmodified plants.

OTHER DEFINITIONS

[0167] Amino acid residues or subunits in a polypeptide sequence are designated herein according to the one-letter code, in which, for example, Q means Gln or Glutamine residue, R means Arg or Arginine residue and D means Asp or Aspartic acid residue.

[0168] Amino acid substitution means the replacement of one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a peptide sequence is an amino acid substitution.

[0169] Nucleotides are designated as follows: one-letter code is used for designating the base of a nucleoside: a is adenine, t is thymine, c is cytosine, and g is guanine. For the degenerated nucleotides, r represents g or a (purine nucleotides), k represents g or t, s represents g or c, w represents a or t, m represents a or c, y represents t or c (pyrimidine nucleotides), d represents g, a or t, v represents g, a or c, b represents g, t or c, h represents a, t or c, and n represents g, a, t or c.

[0170] The term "DNA modifying enzyme" refers to any protein which is capable of modifying the genetic material of a cell, whatever the level of DNA modification (cleavage, covalent interaction, water-mediated interaction . . .). DNA-interacting proteins (e.g., a methylase, a topoisomerase, an integrase, a transposase, or a ligase), transcription activators or repressor, other proteins such as histones, and nucleases are intended to be included in the meaning of "DNA modifying enzyme". When comprised in a TAL effector-DNA modifying enzyme the DNA modifying enzyme is referred as the DNA modifying enzyme domain.

[0171] The term "nuclease" is intended to include exonucleases and endonucleases.

[0172] The term "endonuclease" refers to any wild-type or variant enzyme capable of catalyzing the hydrolysis (cleavage) of bonds between nucleic acids within a DNA or RNA molecule, preferably a DNA molecule. Non-limiting examples of endonucleases include type II restriction endonucleases such as FokI, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, and AlwI. Endonucleases comprise also rare-cutting endonucleases when having typically a polynucleotide recognition site of about 12-45 base pairs (bp) in length, more preferably of 14-45 bp. Rare-cutting endonucleases significantly increase HR by inducing DNA double-strand breaks (DSBs) at a defined locus (Rouet, Smih et al. 1994; Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995; Pingoud and Silva

2007). Rare-cutting endonucleases can for example be a homing endonuclease (Paques and Duchateau 2007), a chimeric Zinc-Finger nuclease (ZFN) resulting from the fusion of engineered zinc-finger domains with the catalytic domain of a restriction enzyme such as FokI (Porteus and Carroll 2005) or a chemical endonuclease (Eisenschmidt, Lanio et al. 2005; Arimondo, Thomas et al. 2006; Simon, Cannata et al. 2008). In chemical endonucleases, a chemical or peptidic cleaver is conjugated either to a polymer of nucleic acids or to another DNA recognizing a specific target sequence, thereby targeting the cleavage activity to a specific sequence. Chemical endonucleases also encompass synthetic nucleases like conjugates of orthophenanthroline, a DNA cleaving molecule, and triplex-forming oligonucleotides (TFOs), known to bind specific DNA sequences (Kalish and Glazer 2005). Such chemical endonucleases are comprised in the term “endonuclease” according to the present document. Examples of such endonuclease include I-Sce I, I-Chu I, I-Cre I I-Csm I PI-Sce I, PI-Tli I, PI-Mtu I I-Ceu I I-Sce II, I-Sce III, HO, PI-Civ I PI-Ctr I, PI-Aae I PI-Bsu I PI-Dha I PI-Dra I, PI-Mav I, PI-Mch I PI-Mfu I, PI-Mfl I PI-Mga I, PI-Mgo I PI-Min I, PI-Mka I, PI-Mle I PI-Mma I, PI-Msh I, PI-Msm I PI-Mth I PI-Mtu I PI-Mxe I PI-Npu I PI-Pfu I, PI-Rma I PI-Spb I PI-Ssp I, PI-Fac I, PI-Mja I PI-Pho I, PI-Tag I, PI-Thy I, PI-Tko I PI-Tsp I I-Msol.

[0173] The endonucleases according to the present document can be part of a Transcription Activator-Like (TAL) effector endonuclease (TALEN).

[0174] By “TALEN” is intended a protein comprising a Transcription Activator-like (TAL) effector binding domain and an endonuclease domain, the fusion of both domains resulting in a “monomeric TALEN”. Some monomeric TALEN can be functional per se and others require dimerization with another monomeric TALEN. The dimerization can result in a homodimeric TALEN when both monomeric TALEN are identical or can result in a heterodimeric TALEN when monomeric TALEN are different. Two monomeric TALEN are different when, for example, their RVDs numbers are different, and/or when the content (i.e amino acid sequence) of at least one RVD is different. By “TAL effector-DNA modifying enzyme” is intended a protein comprising a Transcription Activator-Like effector binding domain and a DNA-modifying enzyme domain.

[0175] By “variant” is intended a “variant” protein, i.e. an protein that does not naturally exist in nature and that is obtained by genetic engineering or by random mutagenesis, i.e. an engineered protein. This variant protein can for example be obtained by substitution of at least one residue in the amino acid sequence of a wild-type, naturally-occurring, protein with a different amino acid. Said substitution(s) can for example be introduced by site-directed mutagenesis and/or by random mutagenesis.

[0176] By “cell” or “cells” is intended any prokaryotic or eukaryotic living cells, cell lines derived from these organisms for in vitro cultures, primary cells from animal or plant origin.

[0177] By “primary cell” or “primary cells” are intended cells taken directly from living tissue (i.e. biopsy material) and established for growth in vitro, that have undergone very few population doublings and are therefore more representative of the main functional components and characteristics of

tissues from which they are derived from, in comparison to continuous tumorigenic or artificially immortalized cell lines. These cells thus represent a more valuable model to the in vivo state to which they refer.

[0178] By “homologous” is intended a sequence with enough identity to another one to lead to homologous recombination between sequences, more particularly having at least 95% identity, preferably 97% identity and more preferably 99%.

[0179] “Identity” refers to sequence identity between two nucleic acid molecules or polypeptides. Identity can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base, then the molecules are identical at that position. A degree of similarity or identity between nucleic acid or amino acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. Various alignment algorithms and/or programs may be used to calculate the identity between two sequences, including FASTA, or BLAST which are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default setting.

[0180] by “mutation” is intended the substitution, deletion, insertion of one or more nucleotides/amino acids in a polynucleotide (cDNA, gene) or a polypeptide sequence. Said mutation can affect the coding sequence of a gene or its regulatory sequence. It may also affect the structure of the genomic sequence or the structure/stability of the encoded mRNA.

[0181] By “gene” is meant the basic unit of heredity, consisting of a segment of DNA arranged in a linear manner along a chromosome, which codes for a specific protein or segment of protein. A gene typically includes a promoter, a 5' untranslated region, one or more coding sequences (exons), optionally introns, a 3' untranslated region. The gene may further comprise a terminator, enhancers and/or silencers.

[0182] The term “gene of interest” refers to any nucleotide sequence encoding a known or putative gene product.

[0183] As used herein, the term “locus” is the specific physical location of a DNA sequence (e.g. of a gene) on a chromosome. The term “locus” usually refers to the specific physical location of a target sequence on a chromosome.

[0184] By “fusion protein” is intended the result of a well-known process in the art consisting in the joining of two or more genes which originally encode for separate proteins, the translation of said “fusion gene” resulting in a single polypeptide with functional properties derived from each of the original proteins.

[0185] By “catalytic domain” is intended the protein domain or module of an enzyme containing the active site of said enzyme; by active site is intended the part of said enzyme at which catalysis of the substrate occurs. Enzymes, but also their catalytic domains, are classified and named according to the reaction they catalyze. The Enzyme Commission number (EC number) is a numerical classification scheme for enzymes, based on the chemical reactions they catalyze (World Wide Web at chem.qmul.ac.uk/iubmb/enzyme/). In the scope of the

present document, any catalytic domain can be used as a partner and be fused to a TAL effector domain to generate a chimeric fusion protein resulting in a TAL effector-DNA modifying enzyme. Non-limiting examples of such catalytic domains can be those of MmE1, EsaSSII, CstMI, NucA, EndA *Escherichia coli*, NucM, EndA *Streptococcus pneumoniae*, SNase *Staphylococcus aureus*, SNase *Staphylococcus hyicus*, SNase *shigella flexneri*, *Bacillus subtilis* yncB, Endodeoxyribonuclease I Enterobacteria phage T7, EndoG bovine, ttSnr DNA mismatch repair protein mutS, cleavage domain of Met-nase.

[0186] The practice of the subject matter disclosed herein will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, *Current Protocols in Molecular Biology* (Ausubel, 2000, Wiley and son Inc, Library of Congress, USA); *Molecular Cloning: A Laboratory Manual*, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); U.S. Pat. No. 4,683,195; *Nucleic Acid Hybridization* (Harries and Higgins eds. 1984); *Transcription and Translation* (Hames and Higgins eds. 1984); *Culture of Animal Cells* (Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells and Enzymes* (IRL Press, 1986); Perbal, *A Practical Guide to Molecular Cloning* (1984); the series, *Methods in Enzymology* (Abelson and Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols. 154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (Goeddel, ed.); *Gene Transfer Vectors For Mammalian Cells* (Miller and Calos eds., 1987, Cold Spring Harbor Laboratory); *Immunochemical Methods in Cell and Molecular Biology* (Mayer and Walker, eds., Academic Press, London, 1987); *Handbook of Experimental Immunology*, Vols. I-IV (Weir and Blackwell, eds., 1986); and *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

[0187] The above written description of the invention provides a manner and process of making and using it such that any person skilled in this art is enabled to make and use the same, this enablement being provided in particular for the subject matter of the appended claims, which make up a part of the original description.

[0188] As used above, the phrases "selected from the group consisting of," "chosen from," and the like include mixtures of the specified materials.

[0189] Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and subranges within a numerical limit or range are specifically included as if explicitly written out.

[0190] The above description is presented to enable a person skilled in the art to make and use the invention, and is provided in the context of a particular application and its requirements. Various modifications to the preferred embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the invention. Thus, this invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

[0191] Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, the invention being further described in the following examples, which do not limit the scope of the invention described in the claims unless otherwise specified.

EXAMPLES

Example 1

A Cipher Governs TAL Effector-DNA Recognition

[0192] To determine whether there is a one-to-one, linear correspondence between RVDs and contiguous nucleotides in the TAL target site, the predicted promoter region (i.e., the 1,000 bp immediately preceding the annotated translational start site) of the known target gene for each of ten TAL effectors was scanned with the TAL effector RVD sequence for alignments that minimized entropy (randomness) in RVD-nucleotide associations. The following formula was used to quantify entropy, where R is the set of RVDs for the effector, D is the set of four nucleotides (A, C, G, T), and $f_{i,j}$ represents the observed frequency with which the i_{th} RVD associates with the j_{th} nucleotide:

$$\sum_{i \in R} \sum_{j \in D} \max_j (f_i) - f_{i,j}$$

[0193] Multiple low entropy sites were present in each promoter. For effector AvrBs3, however, only one mapped to the 54 bp upa20 promoter fragment identified previously as sufficient and necessary for activation, and it coincided with the UPA box common to genes directly activated by AvrBs3 (Kay et al., supra). Also, for effectors PthXo1 and AvrXa27, only one site each overlapped a polymorphism between the activated and non-activated alleles of their respective targets, Os8N3 and Xa27. Across the alignments at these three sites, RVD-nucleotide associations were consistent, so the remaining alignments were selected based on those associations, resulting in exactly one site per TAL effector-target pair (FIG. 1B and Table 2). Each site is preceded by a T (FIG. 1D).

[0194] To assess the specificity conferred by the RVD-nucleotide associations, a weight matrix was first generated based on the frequencies of all RVD-nucleotide associations observed across the ten minimal entropy TAL effector-target site alignments (FIG. 1B). The weight matrix was then used to scan the promoter region, the 1,000 bp preceding the translational start, of each nonredundant gene model in rice, *Oryza sativa* spp. japonica cv. Nipponbare (Osai, Release 6.0, rice.plantbiology.msu.edu) for best matches to the five TAL effectors of the rice pathogen *Xanthomonas oryzae* (AvrXa27, PthXo1, PthXo6, PthXo7, and Tal1c). For AvrXa27, the sequence upstream of Xa27 (GenBank accession AY986492) was included. This upstream sequence is not present in Nipponbare. Observed association frequencies were weighted at 90% and the remaining 10% was distributed equally to frequencies of all possible associations. Alignments were ranked using a weight matrix score (y axis), taken as a negative log of the frequency score derived from the RVD-nucleotide association frequencies in FIG. 1B. Thus, the lower the score, the better the match. For PthXo1, PthXo6, PthXo7, and Tal1c, the experimentally identified target gene was the best

or nearly best match. Better matches were not preceded by a T, were not represented on the microarray used to identify the target, or lacked introns and EST evidence. Scanning the reverse complement promoter sequences yielded no better scoring alignments than the forward sites for the known targets. This result does not imply that TAL effectors bind to the positive strand, but indicates that they function in a forward orientation relative to the positive strand. The known target of the fifth effector, AvrXa27, is the disease resistance gene Xa27 (Gu et al., supra). The poorer rank for this match (5,368) may reflect a calibrated, or recent and sub-optimal host adaptation. Better scoring sites likely comprise genes targeted by AvrXa27 for pathogenesis.

[0195] Using the weight matrix again, ten additional alignments were obtained by scanning all rice promoters with 40 additional *X. oryzae* TAL effectors and retaining the best alignments for which the downstream gene was activated during infection based on public microarray data (PLEXdb.org, accession OS3) (Table 3). As with the initial set, a T precedes each site, and no reverse-strand sites scored better. The RVD-nucleotide association frequencies of the total 20 alignments are shown in FIG. 1C. They constitute a strikingly simple cipher.

[0196] The RVD-nucleotide frequencies in the expanded set of 20 TAL effector nucleotide alignments were used to generate a new weight matrix, and a computational script was written in Python v2.5 (www.python.org). The script can be used to scan any collection of DNA sequences for matches to a particular TAL effector, with a user-definable weight factor for observed vs. unobserved RVD-nucleotide associations. See Moscou and Bogdanove (supra).

[0197] There is some degeneracy in the cipher. Strong associations may represent anchors that account for most of the binding affinity, with weak associations providing a measure of flexibility. Alternatively, neighbor effects may be involved. The latter possibility was investigated by determining the nucleotide association frequencies of every RVD conditioned on the RVD to either side and comparing them to the total observed frequencies—in other words, by sorting the RVD-nucleotide pairings according to the neighbor RVD to the left or right, and comparing the relative frequencies of each pair thus sorted with the overall frequency for that pair. The frequencies of the RVD-nucleotide associations sorted by neighbor did not deviate significantly from the total observed frequencies, suggesting that the associations are context independent.

[0198] Sequences flanking the 20 target sites revealed no conserved nucleotides except the T at −1, but they tend to be C-rich following the site and G-poor throughout (FIG. 1D). With few exceptions, sites begin within 60 bp upstream of the annotated transcriptional start, and none are closer than 87 bp to the translational start (Table 2 and Table 3). Additional rules governing RVD/nucleotide associations are described in Examples 4 and 5.

[0199] Given these results, prediction of TAL effector targets in a genome and construction of targets de novo are now possible. The ability to predict sites will expedite identification of host genes important in disease. The ability to construct targets holds promise for designing durable resistance genes that are responsive to conserved or multiple TAL effectors. Customizing TAL effectors for arbitrary gene activation or targeting of fused proteins for DNA modification also is possible, as described herein.

TABLE 2

Predicted target site features for experimentally identified TAL effector-target pairs						
TAL effector	Source	RVDs	Target gene	TcS	TATA box	TIS
AvrXa27 ¹	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PXO99 ⁴	17	Xa27 (rice)	27	−7	87
AvrBs3 ²	<i>X. campestris</i> pv. <i>vesicatoria</i>	18	Bs3 (pepper)	59	1	123
AvrBs3 ³	<i>X. campestris</i> pv. <i>vesicatoria</i>	18	UPA20 (pepper)	72	1	150
AvrBs3Arep16 ^{4,5}	Modified AvrBs3	14	Bs3-E (pepper)	85	1	136
AvrBs3Arep109 ⁴	Modified AvrBs3	15	Bs3 (pepper)	59	1	123
AvrHah1 ⁶	<i>X. gardneri</i>	14	Bs3 (pepper)	59	1	121
PthXo1 ⁷	<i>X. oryzae</i> pv. <i>oryzae</i> PXO99 ⁴	24	Os8N3 (rice)	79	46	251
PthXo6 ⁸	<i>X. oryzae</i> pv. <i>oryzae</i> PXO99 ⁴	23	OsTFX1 (rice)	31	−780	136
PthXo7 ⁸	<i>X. oryzae</i> pv. <i>oryzae</i> PXO99 ⁴	22	OsTFIIAγ1 (rice)	333	44	469
Tal1c	<i>X. oryzae</i> pv. - <i>oryzicola</i> BLS256	16	OsHEN1 (rice)	10	−265	217

RVDs, repeat-variable diresidues; TcS, annotated transcriptional start site; TIS, translational start site. Locations are relative to the 5' end of the target site.

¹Gu et al., supra

²Kay et al. (2007) *Science* 318: 648

³Römer et al. (2007) *Science* 318: 645

⁴Herbers et al. (1992) *Nature* 356: 172

⁵Römer et al. (2009) *Plant Physiol.*

⁶Schornack et al. (2008) *New Phytologist* 179: 546

⁷Yang et al. (2006) *Proc. Natl. Acad. Sci. USA* 103: 10503

⁸Sugio et al. (2007) *Proc. Natl. Acad. Sci. USA*

TABLE 3

<i>Xanthomonas oryzae</i> TAL effector candidate targets in rice activated during infection.									
Effector	Strain	RVDs	Rice locus	r	TcS	TATA box	TIS	q	Fold change
Tal1c	BLS256	16	OsHen1	1	10	-265	217	0.01	3.3
Tal2c	BLS256	27	Os03g03034	15	-16	-145	143	0.01	5.2
Tal2d	BLS256	16	Os04g49194	9	27	n.p.	102	3.9E-07	29.7
Tal3b	BLS256	18	Os05g27590	42	34	-1	104	3.4E-08	8.5
Tal4a	BLS256	26	Os03g37840	1	152	221	363	2.2E-04	2.6
Tal4b	BLS256	14	Os09g32100	72	68	n.p.	271	8.0E-03	3.6
Tal4c	BLS256	23	Os06g37080	18	31	n.p.	151	2.7E-10	17.1
Tal6	BLS256	20	Os07g47790	16	-15	-70	93	3.6E-02	21.6
PthXo1	PXO99 ^d	24	Os8N3	1	79	46	251	1.0E-08	84.2
PthXo6	PXO99 ^d	23	OsTFX1	2	31	-780	136	3.5E-03	2.8
PthXo7	PXO99 ^d	22	OsTFILAγ1	7	333	44	469	1.6E-06	4.5
Tal9a	PXO99 ^d	20	OsHen1	1	44	-3	93	0.13	8.2
Tal7a/8a	PXO99 ^d	18	Os01g68740	2	32	-197	102	1.8E-01	1.7
Tal7b/8b	PXO99 ^d	20	Os01g40290	57	-2	-276	206	1.8E-01	1.7

RVDs, repeat-variable diresidues; r, rank out of 58,918 gene models scanned, based on the RVD weight matrix score; TcS, annotated transcriptional start site; n.p., not present; TIS, translational start site. Locations are relative to the 5' end of the target site. q values are for a comparison to mock across five time points up to 96 hours after inoculation, replicated four times; fold change given is at 96 hours (PLEXdb, accession OS3).

Example 2

TALENs can Function in Yeast

[0200] Plasmid Construction:

[0201] The protein coding sequence of the TAL effector, AvrBs3, was obtained by digestion from a plasmid with BamHI. A DNA fragment encoding principally the repeat domain was excised with SphI. The amino acid sequence of AvrBs3 can be found under GENBANK Accession No. P14727 and SEQ ID NO:12 (FIG. 3), and the nucleic acid sequence under Accession No. X16130 and SEQ ID NO:13 (FIG. 4). In FIG. 4, the BamHI and SphI sites are in bold and underlined. The AvrBs3 BamHI and SphI fragments were cloned into the nuclease expression vector pDW1789 TAL (FIG. 5) adjacent to sequences encoding the FokI nuclease domain. To clone the AvrBs3 target site into the target reporter plasmid, two complementary DNA oligos, containing two AvrBs3 recognition sites arranged in an inverted orientation with an 18 bp spacer sequence in between, were synthesized with Bg/II and SpeI overhangs at the 5' and 3' ends, respectively. Other reporter plasmids were made that had recognition sites with spacer lengths of 6, 9, 12 and 15 bp. The annealed DNA oligos were cloned into the reporter plasmid, pCP5 (FIG. 6), which was digested with Bg/II and SpeI.

[0202] Yeast Assay:

[0203] The target reporter plasmids were transformed into the yeast strain YPH499 (a MAT α strain), and transformants were selected on synthetic complete medium lacking tryptophan (SC-W). The TALEN expression plasmids were transformed into YPH500 (a MAT α strain); and transformants were plated on SC medium lacking histidine (SC-H). Yeast colonies carrying the target reporter plasmid and colonies carrying the TALEN expression plasmid were cultured overnight at 30° C. in liquid SC-W and SC-H media, respectively. The cultures were adjusted to the same OD₆₀₀, and 200 μl of each were mixed into 200 μl YPD medium. The mixture was incubated at 30° C. for 4 hours to allow the two types of yeast strain to mate. The mixed culture was spun down and resuspended in 5 ml SC-W-H media at 30 C overnight or until the OD₆₀₀ reaches a range of 0.5-1. The cells were harvested and quantitative β-galactosidase assays were performed as described (Townsend et al. (2009) *Nature* 459:442-445).

[0204] Results:

[0205] The TAL-FokI fusion is a site-specific nuclease consisting of the TAL DNA recognition domain and the non-specific FokI DNA cleavage domain. The TAL DNA recognition domain can be engineered to bind different DNA sequences. As described in Example 1 herein, the DNA recognition specificity for TAL effectors, a novel class of DNA binding domain, has been deciphered. In particular, the DNA binding domain of TAL, effectors contain a various number of tandem, 34-amino acid repeats, which can recognize and bind to specific DNA sequences. Amino acid sequences of the repeats are conserved except for two adjacent highly variable residues at positions 12 and 13 of the repeats. These positions together specify individual nucleotides in the DNA binding site, one repeat to one nucleotide. The architecture of the TALENs is illustrated in FIG. 7. The TALENs function as dimers, with each monomer composed of engineered TAL DNA recognition repeats fused to a non-specific cleavage domain from the FokI endonuclease. The DNA recognition repeats can be engineered to bind target DNA sequences within a genome of interest. TAL nuclease monomers bind to one of two DNA half-sites that are separated by a spacer sequence. This spacing allows the FokI monomers to dimerize and create a double-strand DNA break (DSB) in the spacer sequence between the half-sites.

[0206] To explore the potential of the TAL effector DNA recognition domain, experiments were conducted to determine whether native TAL effectors can function as nucleases when fused with the FokI nuclease domain. The yeast-based assay was carried out by using a TAL nuclease expression construct and a target reporter construct. As illustrated in FIG. 5, the backbone of the nuclease expression construct contains a FokI nuclease domain and an N-terminal nuclear localization signal (NLS) under control of the yeast TEF1 promoter. Several restriction sites are located between the FokI nuclease domain and the NLS motif to facilitate cloning of various TAL effectors. The target reporter construct has a disrupted lacZ reporter gene with a 125 bp duplication of coding sequence as shown in FIG. 6. The duplication flanks a URA3 gene and a target sequence (composed of two half sites and a spacer sequence) recognized by TAL DNA binding domains. If the TALEN binds and generates DNA double-strand breaks

(DSBs) at the target site, such breaks, in yeast, are repaired predominantly by homologous recombination between the duplicated lacZ sequences through single strand annealing (Haber (1995) *Bioessays* 17:609). Recombination results in reconstitution of a functional lacZ gene and loss of URA3 (conferring 5-fluoroorotic acid resistance). Relative cleavage activity of TALENs was measured by determining lacZ enzyme activity.

[0207] In these studies, a native TAL effector, AvrBs3, which had a central nuclease repeat region as set forth in SEQ ID NO:3 (FIG. 8) was cloned into the nuclease expression vector, and the AvrBs3 target sites (two binding sites arranged in an inverted orientation) with an 18 bp spacer sequence were cloned into the target reporter vector. The yeast assay was performed using the scheme shown in FIG. 9 and described above. The results showed that the lacZ activity from yeast cells transformed with both the AvrBs3 nuclease plasmid and the target reporter plasmid was significantly higher (15.8-fold higher) than the control yeast cells that contained only the target reporter plasmid (FIG. 10). No activity was observed with nuclease fusions made with only the SphI fragment that encodes predominantly the repeat domain. This indicated that sequences other than the DNA binding domain are required for TALEN activity. Reporter plasmids with spacer lengths of 6 and 9 bp also failed to show activity, indicating that the space between the two binding sites is critical to allow FokI to dimerize. These data indicate that the AvrBs3 TAL nuclease can function as a site-specific nuclease that cleaves its cognate target sequence in yeast.

Example 3

Modular Assembly of TAL Effector Repeats for Customized TALENs

[0208] Complementary oligonucleotides corresponding to the 102 basepairs of each of four individual TAL effector repeats, each specifying a different nucleotide, are synthesized, annealed and cloned into a high copy bacterial cloning vector, individually and in combinations of 2 and 3 repeats in all permutations to yield 4 single, 16 double, and 64 triple repeat modules using standard restriction digestion and ligation techniques (e.g., as illustrated in FIG. 11). The desired TAL effector coding sequence is assembled by introducing the appropriate modules sequentially into a Gateway-ready high copy bacterial cloning vector containing a truncated form of the tal1c gene that lacks the central repeat region except for the characteristic final half repeat. For example, an 18 repeat TAL effector coding sequence can be assembled by sequentially introducing 5 triple modules and 1 double module into the truncated tal1c vector.

Example 4

A System for Modular Assembly of TAL Effector Repeats

[0209] Plasmids and methods were developed for generating custom TAL effector-encoding genes. The functional specificity of TAL effectors is determined by the RVDs in the repeats, as described herein; other polymorphisms in the repeats and elsewhere in the proteins are rare and inconsequential with regard to functional specificity. Thus, custom TAL effector genes were generated by replacing the repeat region of an arbitrary TAL effector gene with repeats containing the desired RVDs. The repeat sequences outside the RVDs

matched a consensus sequence (see below). DNA fragments encoding TAL effector repeats were sequentially assembled into modules encoding one, two, or three repeats, and the modules were cloned into a TAL effector gene from which the original repeats were removed. Each encoded repeat, with the exception of the last (half) repeat, had the sequence LTPDQVVAIASXXGGKQALETVQRLLPVLCQDHG (SEQ ID NO:18; FIG. 12A). The last (half) repeat had the sequence LTPDQVVAIASXXGGKQALES (SEQ ID NO:20; FIG. 12B). In both sequences, "XX" indicates the location of the RVD. The RVDs used in the modular repeats were NI, HD, NN, and NG, which specify binding to A, C, G, and T, respectively. In the experiments described below, the tal1c gene of *Xanthomonas oryzae* pv. *oryzicola* strain BLS256, with its repeats removed, was used as the "backbone" for building custom TAL effector genes.

[0210] The method described herein included five components: (1) generation of single repeat starter plasmids; (2) generation of single repeat module plasmids; (3) generation of multiple repeat modules; (4) generation of a complete set of one-, two-, and three-repeat module plasmids; and (5) assembly of custom TAL effector coding sequences.

[0211] To generate single repeat starter plasmids, the talk gene was digested with MscI and religated to remove the entire repeat region except for the first part of the first repeat and the last part of the last, truncated repeat, resulting in the plasmid designated pCS487 (FIG. 13). The resulting gene encoded the RVD NI and, like most TAL effector genes, contained two SphI sites that flanked the repeat region. The gene contained no XhoI site.

[0212] Next, a translationally silent mutation was introduced into pCS487 to create a unique PspXI site, which encompasses a unique XhoI site centered on codons 19 and 20. The mutation is depicted in FIG. 14, which shows the original and altered nucleotide sequences for codons 18-21 (SEQ ID NO:21 and SEQ ID NO:23, respectively), both of which encode the amino acid sequence ALES (SEQ ID NO:22). The resulting plasmid was designated pCS489.

[0213] By further mutagenesis, three additional constructs were generated with the RVDs HD, NN, and NG, to create the plasmids designated pCS490, pCS491, and pCS492, respectively. The SphI fragment encompassing the modified repeat region was transferred from pCS489, pCS490, pCS491, and pCS492 to the kanamycin resistant plasmid designated pCS488 (FIG. 15), which encoded only the N- and C-terminal portions of tal1c, without the repeat region, in the Gateway entry vector pENTR-D (Invitrogen, Carlsbad, Calif.). This transfer resulted in the single repeat starter plasmids designated pCS493 (FIG. 16), pCS494, pCS495, and pCS496, respectively. The PspXI/XhoI site in the truncated repeat remained unique in these plasmids. The TAL effector gene in pCS488 and each of its derivatives was preceded by Shine-Dalgarno and Kozak sequences for efficient translation in prokaryotes and eukaryotes, respectively.

[0214] Single repeat module plasmids were then constructed. One plasmid was generated for each of the four chosen RVDs (NI, HD, NN, and NG). Each plasmid had a 5' compatible cohesive end that reconstituted a XhoI but not a PspXI site when ligated into a PspXI site, and a 3' compatible cohesive end that reconstituted both a XhoI and a PspXI site. The plasmids were generated by cloning annealed synthetic, complementary oligonucleotides with overhangs (FIG. 17A) into the PspXI/XhoI site of pBluescript SK-, resulting in plasmids designated pCS502 (FIG. 17B), pCS503, pCS504,

and pCS505, respectively. Each plasmid allowed for introduction of additional repeats at the 3' end of the single repeat module at the unique reconstituted PspXI site, or for excision of the repeat module using the reconstituted XhoI sites.

[0215] Additional single repeat modules, one each for NI, HD, NN, and NG, were generated. Each had a 5' compatible cohesive end that did not reconstitute a PspXI or XhoI site when ligated into a PspXI site, a 3' compatible cohesive end that reconstituted both the XhoI and a PspXI site, and a translationally silent nucleotide substitution that destroyed the internal MscI site (FIG. 18A). These modules were generated by annealing synthetic, complementary oligonucleotides with overhangs. Ligating any of these additional single repeat modules into the unique PspXI/XhoI site of a single repeat module plasmid (pCS502, pCS503, pCS504, or pCS505) resulted in no new XhoI site at the 5' junction, but restoration of the unique 3' PspXI/XhoI site, so the resulting plasmids could be linearized for introduction of more additional repeats by cutting with PspXI. Reiteration of this process resulted in modules containing multiple repeats (FIG. 18B). Further, each entire multiple repeat module could be excised using XhoI. Because the MscI site was destroyed in the additional single repeat modules, the MscI site in the initial repeat remained unique, and was useful to check orientation upon subsequent subcloning of the multiple repeat module.

[0216] Additional single repeat modules were cloned iteratively into the single repeat module plasmids to generate, along with the single repeat module plasmids, a complete set of all possible one-, two-, and three-repeat modules, for a total of 84 plasmids designated pCS502 through pCS585 (FIG. 19). Modules containing more than three repeats (e.g., four, five, six, seven, eight, nine, ten, or more than ten repeats) are generated in the same manner.

[0217] A method was then devised to assemble any sequence of repeats into the tal1c "backbone" to generate a custom TAL effector gene. The method included the following steps, which also are depicted in FIG. 20:

[0218] (1) Choose a single repeat starter plasmid with the first desired repeat (pCS493, pCS494, pCS495, or pCS495, encoding RVD NI, HD, NN, or NG, respectively);

[0219] (2) linearize the plasmid with PspXI;

[0220] (3) isolate the module for the next repeat(s) from the appropriate module plasmid (pCS502 through pCS585) using XhoI;

[0221] (4) ligate;

[0222] (5) check orientation by digestion with MscI and confirm sequence from the 3' end using a vector based primer; and

[0223] (6) repeat steps 2-5 until all repeats are assembled.

Example 5

Library of Plasmids for Modular Assembly of TALENs

[0224] Assembly of TALEN repeats as described herein (e.g., using the steps depicted in FIG. 20) results in numerous intermediate plasmids containing increasing numbers of repeats. Each of these plasmids is stored such that a library of plasmids for modular assembly of TALENs (pMATs) is generated. For example, FIGS. 21A and 21B depict the assembly of repeat modules in construction of TAL endonucleases that will target the nucleotide sequences shown. In FIG. 21A, repeat modules from plasmids designated pCS519, pCS524,

pCS537, pCS551, pCS583, and pCS529 are sequentially added to the sequence in the starter plasmid designated pCS493, resulting in plasmids designated pMAT55, pMAT56, pMAT57, pMAT58, pMAT59, and pMAT60. In FIG. 21B, repeat modules from plasmids designated pCS530, pCS533, pCS522, and pCS541 are sequentially added to the sequence in the plasmid designated pMAT1, resulting in plasmids designated pMAT61, pMAT62, pMAT63, and pMAT64.

Example 6

Generation and Testing of Customized TALENs

[0225] The TAL DNA recognition domain was used to create TALENs that recognize and cleave particular DNA targets (FIG. 22A), using the system described in Examples 4 and 5. To assess TALEN function, a yeast assay was adapted in which LacZ activity serves as an indicator of DNA cleavage (Townsend et al., supra). In this assay, a target plasmid and a TALEN expression plasmid are brought together in the same cell by mating. The target plasmid has a lacZ reporter gene with a 125-bp duplication of coding sequence. The duplication flanks a target site recognized by a given TALEN. When a double-strand DNA break occurs at the target site, it is repaired through single-strand annealing between the duplicated sequences, which creates a functional lacZ gene whose expression can be measured using standard β -galactosidase assays that provide a quantifiable readout (FIG. 22A). This assay has been demonstrated to be a good predictor of the ability of a ZFN to create chromosomal mutations by NHEJ or to stimulate homologous recombination for gene editing in higher eukaryotes (Townsend et al., supra; and Zhang et al. (2010) *Proc. Natl. Acad. Sci. USA* 107:12028-12033).

[0226] Two well characterized TAL effectors were used—AvrBs3 from the pepper pathogen *Xanthomonas campestris* pv. *vesicatoria* and PthXo1 from the rice pathogen *X. oryzae* pv. *oryzae* (Bonas et al. (1989) *Mol. Gen. Genet.* 218:127-136; and Yang et al. (2006) *Proc. Natl. Acad. Sci. USA* 103:10503-10508). The amino acid sequence of AvrBs3 can be found under GENBANK Accession No. P14727 and SEQ ID NO:12 (FIG. 3), and the nucleic acid sequence under Accession No. X16130 and SEQ ID NO:13 (FIG. 4). The amino acid sequence of PthXo1 can be found under GENBANK Accession No. ACD58243 and SEQ ID NO:31 (FIG. 23), and the nucleic acid sequence under Accession No. CP000967, gene ID 6305128, and SEQ ID NO:32 (FIG. 24). The amino acid sequence of PthXo1 under GENBANK Accession No. ACD58243 is truncated at the N-terminus due to a misannotation of the start codon. The complete sequence is presented in FIG. 23.

[0227] The repeat domains of both AvrBs3 and PthXo1 are encoded entirely within a conserved SphI fragment (FIGS. 4 and 24). Both TAL effector-encoding genes also have a BamHI restriction fragment that encompasses the coding sequence for the repeat domain and 287 amino acids prior and 231 amino acids after (FIGS. 4 and 24; see, also, FIG. 22A). Absent from the BamHI fragment is the TAL effector transcriptional activation domain. Both the SphI fragments and the BamHI fragments were fused to a DNA fragment encoding FokI that is present in the nuclease expression vector pFZ85 (FIG. 25). The fusion proteins between FokI nuclease and the BamHI fragments encoded by AvrBs3 and PthXo1 are given in FIGS. 26 and 27; SEQ ID NOS:33 and 34.

[0228] The FokI monomers must dimerize in order to cleave, but the appropriate spacer length between the two DNA recognition sites was unclear. For ZFNs, in which the zinc finger array is separated from FokI by a 4-7 amino acid linker, the typical spacer between the two recognition sites is 5-7 bp (Handel et al. (2009) *Mol. Ther.* 17:104-111). Since, for example, 235 amino acids separate the repeat domain from FokI in the BamHI TALEN constructs used herein, a variety of spacer lengths for both the BamHI and SphI constructs (6, 9, 12, 15, and 18 bp) were used. As a positive control, a well-characterized zinc finger nuclease with a DNA binding domain derived from the mouse transcription factor Zif268 (Porteus and Baltimore (2003) *Science* 300:763) was used. As negative controls, the TAL effector domains were fused to a catalytically inactive FokI variant or tested against non-cognate DNA targets.

[0229] Haploid cell types containing either TALEN expression or target plasmid in 200 μ l of overnight culture were mated in YPD medium at 30° C. After 4 hours, the YPD medium was replaced with 5 ml of selective medium and incubated overnight at 30° C. Mated cultures were lysed, ONPG substrate added, and absorbance read at 415 nm using a 96-well plate reader (Townsend et al., supra). β -galactosidase levels were calculated as a function of substrate cleavage velocity. The results obtained with target reporter constructs that had a 15 bp spacer separating the two recognition sites are shown in FIG. 22B. All nuclease expression constructs derived from the SphI fragment, which encoded principally the repeat array, failed to show activity, indicating that amino acid sequences in addition to those in the repeat array are required for function (FIG. 22B). Robust activity, however, was observed for both the AvrBs3 and the PthXo1 TALENs derived from the BamHI fragment (FIG. 22B). The activity of the PthXo1 TALEN approximated that of the ZFN positive control. The activity required the functional FokI domain and was specific for the DNA target recognized by a given TALEN.

[0230] Experiments also were conducted to test various distances between the TAL effector binding sites (11 length variants between 12 and 30 bp), in order to identify spacer lengths that enable FokI to dimerize most efficiently (FIG. 28A). Both enzymes showed two spacer length optima—one at 15 bp and the other at either 21 bp (AvrBs3) or 24 bp (PthXo1). For PthXo1, activity was observed for all tested spacer lengths 13 bp and longer. Some spacer lengths for AvrBs3 showed no activity, however, suggesting that spacer length is critical for certain TALENs.

[0231] The above experiments tested activity of homodimeric TALENs, which bind two identical recognition sequences placed in opposition on either side of the spacer. Since such palindromic sites are unlikely to occur naturally in genomic targets, experiments were conducted to test whether TALENs could function as heterodimers. AvrBs3 and PthXo1 recognition sites were placed in head to tail orientation on either side of a 15 bp spacer. Activity of AvrBs3 and PthXo1 TALENs individually and Zif268 on their respective targets was measured as controls. As a negative control, a yeast culture with only the target site plasmid for the heterodimeric site was assayed for LacZ activity. The resulting activity of the heterodimeric TALEN approximated an average of the activities observed with the two homodimeric enzymes (FIG. 28B). To test whether repeat domains can be assembled to target TALENs to arbitrary chromosomal sequences, two genes were chosen that were previously targeted for mutagen-

esis with ZFNs—ADH1 from *Arabidopsis* and gridlock from zebrafish (Foley et al. (2009) *PLoS One* 4:e4348; and Zhang et al., supra). A search was conducted for 12-13 bp sequences in the coding regions that were preceded by a 5' T and with a nucleotide composition similar to that of TAL effector binding sites identified by Moscou and Bogdanove (supra). In ADH1 and gridlock, such sites occurred on average every 7-9 bp. Four 12 bp sites were selected in ADH1 (at positions 360, 408, 928, and 975 of the chromosomal gene sequence) and one 13 bp site in gridlock (at position 2356 of the chromosomal gene sequence; FIG. 29A). TAL effector repeat domains were constructed to recognize these targets, using the most abundant RVDs from native TAL effectors (NI for A, HD for C, NN for G, and NG for T). To construct custom TALENs, repeats with these RVDs were synthesized individually and assembled into modules of one, two, or three repeats as described in Examples 4 and 5. These modules were ligated sequentially into a derivative of the talk gene (Moscou and Bogdanove, supra) from which the original repeats had been removed, and BamHI fragments from these engineered TAL effectors were fused to sequences encoding the catalytic domain of FokI in pFZ85 (FIG. 25). Five custom TALENs targeted to ADH1 from *Arabidopsis* and the zebrafish gridlock gene were created.

[0232] The resulting custom TALENs were tested in the yeast assay as homodimeric TALENs (that is, the identical DNA binding site was duplicated in inverse orientation on either side of a 16-18 bp spacer), although it is noted that heterodimeric TALENs would need to be constructed to direct cleavage at naturally occurring DNA targets. Spacer lengths were chosen based on the distance closest to 15 bp from the 3' end of the next neighboring (and opposing) candidate site. Sixteen by spacers were used for ADH1-360-12, ADH1-408-12r, and 18 bp spacers for ADH1-928-12, ADH1-975-12r, and gridlock-2356-13r. The yeast assay was performed as described above.

[0233] Robust nuclease activity was observed for the ADH1-360-12 and gridlock-2356-13r TALEN (FIG. 29B). The ADH1-928-12 TALEN had modest activity that was nonetheless significantly above the negative controls. For each TALEN that gave positive results, nuclease activity was specific to the cognate target. These results indicate that novel, functional TALENs can be created by assembly of customized repeat domains.

Example 7

Naturally Occurring Target and TAL Effector Pairs Show Overall and Positional Bias in Nucleotide and RVD Composition

[0234] The 20 paired targets and TAL effectors analyzed by Moscou and Bogdanove (supra) were evaluated for overall composition bias and for positional effects on nucleotide or RVD frequencies. It was observed that sites (on the positive strand) were generally A- and C-rich, and G-poor. The average percent A was 31 $\pm$ 16% (1 standard deviation). The average percent C was 37 $\pm$ 13%. The average percent G was 9 $\pm$ 8%, and the average percent T was 22 $\pm$ 10%. Since the alignments vary in length, the analysis of positional effects was restricted to the five positions on each end. Strikingly, bias in the target sequences was apparent for A and against T at positions 1 and 3, and for T at position N and possibly 2. G was particularly rare at position N-1. This bias was reflected by matching RVDs in the effectors, with NI being most com-

mon at positions 1 and 3, no NG at position 1, nearly always NG at position N, and rarely NN at position N-1 (FIG. 30).

Example 8

Method and Reagents for Rapid Assembly and Cloning of Custom TAL Effector Repeat Arrays

[0235] The Golden Gate cloning method [Engler et al. (2008), supra; and Engler et al. (2009), supra] employs the ability of Type IIS restriction endonucleases (e.g. BsaI) to cut outside their recognition sites to create custom overhangs for ordered ligation of multiple DNA fragments simultaneously. Using this method, several DNA fragments can be fused into an array in a specific order and cloned into a desired destination vector in a single reaction (FIG. 31).

[0236] A method and reagents for assembling custom TAL effector repeat encoding arrays were developed based on the Golden Gate system. When BsaI sites are positioned on either side of a TAL effector repeat coding sequence, cleavage releases a repeat fragment flanked by 4-bp overhangs. Because the cleavage site is not sequence-specific, by staggering, repeat clones can be released with ordered, complementary overhangs (sticky ends), enabling the ordered assembly of multi-repeat arrays.

[0237] A library of 58 plasmids (FIGS. 32A and 32B) was generated to allow the simultaneous assembly of up to 10 repeat units into "subarrays," followed by simultaneous assembly of one, two, or three of these subarrays together with a final truncated repeat into a complete, custom array. Ten staggered sets of four fragments, each fragment in a set encoding a repeat module with a different one of the four most common RVDs, HD, NG, NI, and NN, were synthesized and cloned into a vector carrying the tetracycline resistance gene, for a total of 40 plasmids. Four more fragments that encoded the terminal truncated TAL effector repeat of 20 amino acids, each fragment encoding a different one of the four most common RVDs, were synthesized and cloned into a different vector carrying the spectinomycin resistance gene to yield four more plasmids, designated as "last repeat plasmids," (FIG. 32A). All fragments in the staggered sets are flanked by BsaI sites in the vector so that cleavage with BsaI releases the fragments with different sticky ends that allow for assembly in the appropriate order; that is, i.e. the overhang at the 3' end of a fragment for repeat module 1 is complementary only to the overhang at the 5' end of the fragment for repeat module 2, the overhang at the 3' end of repeat module 2 is complementary only to the overhang at the 5' end of repeat module 3, and so on. The fragments in the last repeat plasmids are flanked by sites for a different Type IIS restriction endonuclease, Esp3I. Fourteen additional plasmids, described following, were constructed as destination vectors to receive assembled subarrays.

[0238] The first destination vector, plasmid pFUS_A was constructed to receive the first subarray of 10 repeats to be assembled into a final array of 21 or fewer repeats (counting the final, truncated repeat). pFUS_A was constructed such that cleavage by BsaI creates an overhang on one side complementary to the overhang at the 5' end of the first repeat module and an overhang at the other side complementary to the overhang at the 3' end of the 10th repeat module. To receive a second subarray of 10 or fewer repeats to be assembled into such a final array, destination vector plasmids pFUS_B1, pFUS_B2, pFUS_B3, pFUS_B4, pFUS_B5, pFUS_B6, pFUS_B7, pFUS_B8, pFUS_B9, and pFUS_B10

were constructed that when cleaved by BsaI have overhangs respectively complementary to the overhang at the 5' end of the first repeat module and the 3' end of the repeat module for the corresponding numbered position (e.g., the pFUS_B6 overhang for the 3' end of the subarray matches the overhang of the four repeat module fragments for position 6). Arrays cloned in pFUS_A and the pFUS_B series of plasmids are flanked by Esp3I sites in the vector and when released by digestion with Esp3I the arrays have unique complementary overhangs that allow for them to be ligated in order along with a final truncated repeat fragment into destination vector pTAL, which encodes a TALEN missing the repeat region. pTAL was constructed so that cleavage with Esp3I allows insertion of the repeat array at the correct location and in the correct orientation by virtue of an overhang at one end that is complementary to the overhang at the 5' end of the first ten repeat subarray and an overhang at the other end complementary to the overhang at the 3' end of the final truncated repeat fragment (FIG. 33).

[0239] The final two destination vector plasmids, pFUS_A30A and pFUS_A30B were constructed to receive the first and second ten repeat subarrays to be assembled into a final array of 22-31 repeats. pFUS_A30A and pFUS_A30B were constructed such that digestion with Esp3I releases the arrays with the appropriate complementary overhangs such that the arrays can be ligated in order along with a third array from a pFUS_B vector and a final truncated repeat fragment from a last repeat plasmid, released similarly by digestion with Esp3I, into pTAL (FIG. 32).

[0240] All destination vectors have the LacZ gene cloned in between the Type IIS restriction endonuclease sites, allowing for blue-white screening for recombinants. Except for pTAL, which carries a gene for ampicillin resistance, all the destination vectors carry a gene for spectinomycin resistance.

[0241] To rapidly construct a custom TAL effector repeat array using these reagents, the following method was established. In the first step, the appropriate individual RVD module plasmids for the necessary subarrays of ten or fewer repeats are mixed together with the appropriate destination vector in one tube. T4 DNA ligase and BsaI endonuclease are added and the reaction is incubated in a PCR machine for 10 cycles of 5 minutes at 37° C. and 10 minutes at 16° C., the respective optimal temperatures for the two enzymes. The reaction mixture is then treated with the PLASMID-SAFET™ nuclease to hydrolyze all linear dsDNA fragments in order to prevent cloning of shorter, incomplete arrays by in vivo recombination, and then the mixture is used to transform chemically competent *E. coli* cells. The resulting recombinant plasmids are isolated and the correct constructs confirmed. Then, in the second step, the confirmed plasmids from the first step are mixed together with the appropriate last repeat plasmid and pTAL, and the digestion and ligation reaction cycle carried out as in the first step. Finally, the reaction products are introduced into *E. coli*, and the full length, final array construct is isolated and confirmed. The protocol can be completed by one person within a week's time.

[0242] Expression constructs for TALENS 85, 102 and 117 in Table 4A, as well as TALENS HPRT-3254-17 and HPRT-3286-20r, described in Example 14 below, were made using the method and reagents described in this example.

[0243] Repeat arrays cloned in pTAL are subcloned readily into other TAL effector gene contexts using the conserved SphI restriction endonuclease sites that flank the repeat region.

Example 9

Custom TALEN Data Show Initial Support for “Rules” and a Correlation Between RVD Number and Activity

[0244] Example 6 describes experiments conducted to engineer the TALEN DNA binding domain so that it can recognize unique DNA sequences. As described, these custom TALENs recognized sites in the *Arabidopsis* ADH1 and zebrafish gridlock genes. Additional custom TAL effector DNA binding domains were engineered to recognize not only sites in these genes, but also in the TT4 gene from *Arabidopsis*, and telomerase from zebrafish (Foley et al., supra; and Zhang et al., supra). These custom TALENs were made using the methods described in Examples 3, 4 and 8. In engineering the custom TALENs, the observed compositional and positional biases were adopted as design principles or “rules.” First, a search was conducted for sequences in the coding regions that were preceded by a 5' T and at least 15 bp in length, and that had a nucleotide composition consistent with the averages noted above. Specifically, only those sites with 0-63% A, 11-63% C, 0-25% G, and 2-42% T were selected. Such sites occurred on average every 7-9 bp. Sites were then selected that conformed to the observed positional biases described above. From this set, two pairs of binding sites in each gene were identified that were 15-19 bp in length and separated by 15-18 bp, so that binding of the engineered TALENs would allow FokI to dimerize. The modular assembly methods (Examples 3 and 4) generated partial length constructs.

[0245] In total, 21 intermediate and full length TALENs designed to target 16 nucleotide sequences, each with an array of nine repeats or longer. The amino acid sequences of these TALENs are provided in FIGS. 34A-34U (SEQ ID NOS:35-55). These 21 TALENs were tested for their ability to cleave DNA using the yeast assay described in Examples 2 and 6. Activity data are shown in FIG. 35 and summarized in Table 4A.

[0246] Some of the intermediate, partial length TALENs correspond to targets that break the rules for nucleotide com-

position and terminal T. Table 4A shows length, conformity to these two rules, and activity relative to that of ZFN268 for each TALEN. The results reveal a general trend that increasing the length of the RVD array increases activity of the resulting TALEN. This suggests that there is a minimal number of RVDs that are needed before a DNA target can be recognized in vivo. Further, conformity to the rules appears to be important. Of the six TALENs showing no detectable activity, two violated the target composition rule, two did not end in NG, and another broke both rules (one obeyed both rules). Three of the eight TALENs with activity less than 25% of ZFN268 violated one of the rules, and one of four TALENs with activity 25-50% of ZFN268 did not have an RVD sequence ending in NG. It is noted that TALENs with activity 50% or greater than that of ZFN268 obeyed all the rules, and for TALENs of the same length, rule breakers generally had less activity than obedient arrays. Consistent with the overall trend regarding length, even for intermediates that broke no rules, the corresponding full length TALENs had higher activity (Table 4A and FIG. 35). Variation in spacer length due to TALEN length differences on the same target may have contributed to this observation, but some range of spacer lengths is tolerated (Christian et al., supra).

[0247] Some complexities in the data were apparent. For example, activity varied among obedient TALENs of the same length, some short arrays had moderately high activity, and some long arrays that were obedient had little or no activity (Table 4B). Nonetheless, the results provided support for the conclusions that 1) generally a greater number of repeats results in greater activity, and 2) conformity to composition and positional bias rules is important for activity. Therefore, the following design principles were derived.

[0248] TAL effector binding sites are designed to be a minimum of 15 bases long and oriented from 5' to 3' with a T immediately preceding the site at the 5' end.

[0249] A site may not have a T in the first (5') position or an A in the second position.

[0250] A site must end in T (3'), and may not have a G at the next to last position.

[0251] The base composition of the site must fall within specified ranges (average±two standard deviations): A 0-63%, C 11-63%, G 0-25%, and T 2-42%.

TABLE 4A

Activity, conformity to rules, and length of TALENs tested in the yeast assay.							
Gene	TALEN	Names from Christian et al. (supra)	RVDs	Activity	% GATC	Ends in NG	RVD sequence ¹
telomerase	124		9	–	N	Y	HD NN NN NG NG NG NN HD NG
gridlock	105		10	+	N	N	NI HD HD HD HD NG HD NG HD HD
ADH1	58	ADH1-360-12	12	++	Y	N	NI NG HD NI NI NN NI NG NG HD NG HD
ADH1	63	ADH1-408-12r	12	–	Y	N	HD HD HD NI NN NI NI NN NG NI NI NI
ADH1	68	ADH1-928-12	12	+	Y	Y	HD HD NN NN NI NG NN HD NG HD HD NG
ADH1	73	ADH1-975-12r	12	–	N	N	NI NN NI HD NI NI NI HD HD NI HD NI
TT4	89		12	–	Y	N	NN NN HD NI HD NG NN HD NG NI NI HD
gridlock	106	gridlock-2356-13r	13	++	Y	Y	NI HD HD HD HD NG HD NG HD HD NN HD NG
ADH1	64		15	+	Y	Y	HD HD HD NI NN NI NI NN NG NI NI NI HD NI NG
ADH1	69		15	+++	Y	Y	ND ND NN NN NI NG NN HD NG HD HD NG HD NG NG
ADH1	74		15	++	Y	Y	NI NN NI HD NI NI NI HD HD NI HD NI NI HD NG
TT4	90		15	–	Y	Y	NN NN HD NI HD NG NN HD NG NI NI HD HD HD NG
telomerase	121		15	+	Y	Y	HD NG NG NN NG HD HD NN HD NI NG NN NI NG NG

TABLE 4A-continued

Activity, conformity to rules, and length of TALENs tested in the yeast assay.						
Gene	Names from Christian et al. TALEN (supra)	RVDs	Activity	% GATC	Ends in NG	RVD sequence ¹
telomerase	126	15	—	N	Y	HD NN NN NG NG NG NN HD NG NI NG HD NN NG NG
gridlock	107	16	++++	Y	Y	NI HD HD HD HD NG HD NG HD HD NN HD NG NG HD NG
gridlock	117	16	++	Y	Y	HD HD HD NN NN NI NI NN HD HD NN NI HD NN HD NG
telomerase	131	16	+	Y	Y	NI NG NG HD HD HD HD NI HD NN NI NN HD NG HD NG
telomerase	136	17	+	N	Y	NI NN NI HD NI NN NN NI NI NN NG NN NN NI NN HD NG
ADH1	60	18	+++++	Y	Y	NI NG HD NI NI NN NI HG NG HD NG HD NG NG HD NI HD NG
TT4	85	18	+	Y	Y	NI HD NG HD HD NN HD HD NG NN NI NI NN HD NI HD NI NG
gridlock	102	18	+	Y	N	NN NN HD NG HD NI HD HD NG NI HD NI NI HD NN NI HD NI

¹Target sequences tested consist of inverted repeats of the corresponding nucleotide sequence, where HD, NG, NI, and NN correspond to C, T, A, and G, respectively, separated by a spacer sequence of 16-18 bp.

TABLE 4B

Excerpt of Table 4A, sorted by activity level			
RVDs	Activity	% GATC	Ends in NG
9	—	n	y
12	—	y	n
12	—	n	n
12	—	y	n
15	—	y	y
15	—	n	y
10	+	n	n
12	+	y	y
15	+	y	y
15	+	y	y
16	+	y	Y
17	+	n	y
18	+	y	y
18	+	y	n
12	++	y	n
13	++	y	y
15	++	y	y
16	++	y	y
15	+++	y	y
16	++++	y	y
18	+++++	y	y

Example 10

Heterodimeric TALEN Pairs Cleave their Intended Naturally Occurring Target Sequences in the Yeast Assay

[0252] The data in Examples 2, 6 and 9 demonstrate that custom TALENs can be engineered to recognize novel target DNA sequences. The yeast activity data for the custom TALENs was gathered using individual TALEN monomers that recognized a homodimeric target site. That is, the target sequence of the TALEN was duplicated in inverse orientation on either side of a 15-18 bp spacer. Cleavage of endogenous chromosomal sequences, however, generally would require that two different custom TALENs recognize two different sequences on either side of a spacer. As described in Example 6, this ability was demonstrated for the AvrBs3 and PthXo1 TALENs together using a corresponding chimeric target site in the yeast assay. We tested whether two different custom TALENs could recognize and cleave a naturally occurring DNA sequence. Using the yeast assay described in Example 2, custom TALENs designed to cleave two different target sequences in the *Arabidopsis* ADH1 gene were assayed for

activity on these targets. The DNA sequences of the target sites and the corresponding TALENs are shown in FIG. 36A. The amino acid sequences of the TALENs are provided in FIG. 34. The beta-galactosidase activity obtained in the yeast assay is plotted in the graph shown in FIG. 36B. The activity of the TALENs on their naturally occurring target sequence was significantly above the negative controls, indicating that TALENs can be engineered to recognize and cleave endogenous target DNA sequences.

Example 11

TALENs Cleave Native Genes in *Arabidopsis* and Introduce Mutations by Imprecise Non-Homologous End-Joining

[0253] One of the active TALEN pairs designed to recognize a target sequence in the *Arabidopsis* ADH1 gene was tested to determine whether it can bind, cleave and mutate chromosomal DNA. Each of the individual ADH1 TALENs comprising this pair (pTALENs 69 and 74) was cloned into the plant expression vector pFZ14, which places the TALENs under the control of the constitutive 35S promoter (Zhang et al., supra). The resulting constructs were then introduced into *Arabidopsis* protoplasts by electroporation. After 48 hours, genomic DNA was isolated and digested with Tth1111. A Tth1111 cleavage site is located in the spacer sequence between the two TALEN recognition sites (FIG. 37A). Cleavage of the chromosomal DNA by the TALEN would be expected to introduce mutations by imprecise non-homologous end joining (NHEJ), which would result in failure to cleave by Tth1111. A 375 bp fragment encompassing the TALEN recognition site was then PCR amplified. The PCR product was digested again with Tth1111 to remove most of the remaining genomic DNA that was not modified by TALEN-mediated NHEJ. The digestion products were then run on an agarose gel. An uncleaved PCR product was observed, and such uncleaved PCR products are diagnostic of nuclease activity (in this case TALEN activity) at the endogenous target sequence (Zhang et al., supra). The uncut DNA was cloned and analyzed by DNA sequencing. The sequencing of nine independent clones revealed that six carried mutations introduced by NHEJ (FIG. 37B). Thus, TALENs cleave endogenous chromosomal loci and introduce DNA double strand breaks and mutations.

Example 12

Enhancing Targeting Capacity

[0254] At the core of the TAL effector DNA cipher, the four most common RVDs each have apparent one-to-one specific-

ity for the four nucleotides, based on association frequencies. This is markedly so for HD, NG, and NI, but less so for NN (FIG. 1C). NN associates most frequently with G, but almost as commonly with A, and sometimes with C or T. For a randomly assembled TAL effector with NN at four locations in a 13 RVD sequence, having G at all corresponding positions in an artificial target gave the best activity (Boch et al. (2009) *Science* 326:1509-1512). A reduced but did not abolish activity, and C and T eliminated detectable activity. A drastic loss of activity was observed when C, T, or A was substituted for G at just the first position in the binding site for the 24 RVD effector PthXo1, which is an NN (Romer et al. (2010) *New Phytol.* 187:1048-1057). This was in contrast, however, to the observation that the much shorter AvrHah1 (14 RVDs) begins with an NN that aligns with A, and the 23 RVD effector PthXo6 has three NNs in a row at positions 4-6 that each align with A, yet both of these proteins are highly active (see, Schornack et al. (2008) *New Phytol.* 179:546-556; and Romer et al., supra). Thus the specificity of NN for G appears to be generally weak and can vary with context.

[0255] The observed invariance of the thymine immediately preceding TAL effector target sites is a requirement for several effectors [Boch et al., supra; Romer et al., supra; and Romer et al. (2009) *Plant Physiol.* 150:1697-1712]. The amino acid sequence immediately preceding the repeat region in TAL effectors, which is highly conserved (FIG. 38A), shares significant similarity with the repeat, both in amino acid sequence and in predicted secondary structure (FIG. 38B and Bodganove et al. (2010) *Curr. Opin. Plant Biol.* 13:394-401). It was hypothesized that this sequence, termed the “0th” repeat, is the basis for the requirement for T at position -1 of the binding site, and that residues in the RVD-analogous position (FIG. 38B) specify the nucleotide.

[0256] Based on these findings, it was hypothesized that by incorporating repeats with high specificity for G, and by relaxing the requirement for T at -1, targeting capacity for engineered TAL effector proteins can be enhanced. Experiments were initiated to test novel and rare RVDs for more robust specificity for G than NN displays, and to replace the RVD-analogous residues of the 0th repeat with common RVDs.

[0257] Novel and Rare RVDs for Robust Specificity for G:

[0258] The modules disclosed above (see, e.g., Example 4) used four particular RVDs (NI, HD, NN, and NG) to specify binding to the four nucleotide bases (A, C, G, and T, respectively). Repeats containing other RVDs also may be useful, and may have increased specificity and/or affinity for the four bases as compared to NI, HD, NN, and NG. Toward improving specificity for G, several repeats encoding novel and rare RVDs were constructed. The rare RVDs NK, HN, and NA associated with G, suggesting that N may be important as one or the other of the residues (FIG. 1C). Thus, a broad set of derivatives encoding repeats having the RVDs shown in Table 5 were constructed. The left column lists RVDs having a polar amino acid (R, K, D, E, Q, H, S, T, or Y) at position 12 and N at position 13. The right columns list combinations of N in the first position with any of 17 other amino acids (G, L, V, R, K, D, E, Q, H, T, M, C, P, Y, W, or F) in the second position of the RVD. To account for the possibility of greater specificity without N, repeats also were made with a polar amino acid (R, K, D, E, Q, H, S, T, or Y) at position 12 and a gap (*) at position 13 (middle column).

[0259] Novel artificial RVDs are tested for function in a quantitative reporter gene based assay for transcriptional acti-

vation activity of TAL effectors, such as a GUS or dual luciferase reporter based, *Agrobacterium*-mediated transient expression assay in *Nicotiana benthamiana*, or in the lacZ reporter based TALEN assay in *Saccharomyces cerevisiae*, described above (see, e.g., Example 2) Repeat modules containing RVDs to be tested are incorporated into a TAL effector or TALEN with measurable and sub-saturation levels of activity, and the resulting proteins are tested for differences in activity on a set of DNA targets with integrated permutations of all four nucleotides at corresponding positions. In particular, beginning with the PthXo1 variant(s) minimally active in the in planta and yeast assays and responsive to mismatches at three added repeats, TALENs containing each of the novel and rare repeats (in homomeric threes) are tested in vivo against targets with G at each of the corresponding positions. For any that show increased activity, the assays are repeated with targets permuted to the other nucleotides at those positions, to ascertain specificity.

TABLE 5

RVDs to be tested ^a			
Polar + N	Polar*	N + all	
RN	R*	NG	NH
KN	K*	NA	NT
DN	D*	NL	NM
EN	E*	NV	NC
QN	Q*	NR	NP
HN	H*	NK	NY
SN	S*	ND	NW
TN	T*	NE	NF
YN	Y*	NQ	

^aNI, NG, and NS nt association frequencies are known. An asterisk represents a gap corresponding to the 2nd position in the RVD (i.e., the 13th position of the consensus repeat sequence).

[0260] Common RVD Substitutions for the RVD-Analogous Position of the 0th Repeat to Relax Specificity of Tat Position -1:

[0261] Secondary structure predictions and alignment of the 0th repeat and repeat consensus sequences suggested that positions occupied by KR* (asterisk denotes a gap) in the 0th repeat were analogous to the RVD and were therefore the residues that specify the T at -1. Variants of PthXo1 with substitutions of HD, NG, NI, and NN for KR and separately for R* were constructed in the Tal1c “backbone” construct described above. Activities of these variants are compared to the wild type effector in the in planta and yeast assays using targets with corresponding nucleotides at position -1, namely, C, T, A, and G, respectively. Additional variants of PthXo1 are constructed that have S, the residue at position 11 of the consensus repeat sequence, substituted for the K at position 11 of the 0th repeat. And other variants are constructed that have this substitution combined with a substitution of K, the residue at position 16 of the consensus repeat sequence, for the V at position 15 of the 0th repeat (Table 6). A proximal TATA box for TAL effector activity may be included. In addition, PthXo1 is useful for this experiment because unlike AvrBs3, for which the T at -1 appears to be part of a TATA box, the TATA box closest to the PthXo1 binding site is 46 bp downstream and would not be perturbed by modifications at -1.

[0262] If the above modifications do not result in enhanced targeting for G or increased ability to target sequences preceded by nucleotides other than T, then a more comprehensive set of artificial RVDs are tested for G specificity, and substitutions other than the common RVDs are tested for the 0th repeat.

TABLE 6

0 th repeat constructs to be made and tested for specificity for targets with A, C, G, or T at the -1 position				
Native 0 th repeat sequence (specifies T at -1)	Substitution specifying T	Substitution specifying A	Substitution specifying C	Substitution specifying G
. . . KIA*	. . . KIA* NGGGV . . .	. . . KIA* NIGGV . . .	. . . KIA* HDGGV . . .	. . . KIA* NNGGV . . .
KRGGV . . . (74) [†]	(75)	(76)	(77)	(78)
. . . KIA*	. . . KIA <u>S</u> NGGGV . . .	. . . KIA <u>S</u> NIGGV . . .	. . . KIA <u>S</u> HDGGV . . .	. . . KIA <u>S</u> NNGGV . . .
KRGGV . . . (79)	(80)	(81)	(82)	(83)
. . . KIAKR*	. . . KIAK NGGGV . . .	. . . KIAK NIGGV . . .	. . . KIAK HDGGV . . .	. . . KIAK NNGGV . . .
GGV . . . (84)	(85)	(86)	(87)	(88)
. . . KIA*	. . . KIA <u>S</u> NGGGK . . .	. . . KIA <u>S</u> NIGGK . . .	. . . KIA <u>S</u> HDGGK . . .	. . . KIA <u>S</u> NNGGK . . .
KRGGV . . . (89)	(90)	(91)	(92)	(93)

Candidate and substituted RVDs are bold.

Other substitutions or modifications are underlined.

Asterisks denote a gap relative to the consensus repeat sequence.

[†]SEQ ID NO:

Example 13

Novel Predicted Nucleotide Specific RVDs

[0263] It was observed that when the RVDs listed in Tables 1A and 1B were grouped by the second amino acid residue in the RVD (i.e., the 13th in the overall repeat), there was a near perfect correlation of that amino acid with the nucleotide(s) specified by the RVD, irrespective of the amino acid at the first position of the RVD (Table 7). Thus, RVDs ending in a gap (denoted by an asterisk) specify C or T, or T; RVDs ending in D specify C; RVDs ending in G specify T; and RVDs ending in N specify G or A, or G. It also was observed that amino acids at position 1 of the RVD were either H, I, N, S, or Y. These observations suggested that RVD specificity is determined by the residue in the second position, independent of whether the residue at the first position is H, I, N, S, or Y. Therefore, specificities were predicted for several novel (i.e., yet unobserved) RVDs that combine residues observed at the second position with residues H, I, S, N, or Y at the first position. Thus, I*, S*, and Y* were predicted to specify C or T, or T; ID, SD, and YD were predicted to specify C; SG was predicted to specify T; and IN and YN were predicted to specify G or A, or G. Also, although there was only one instance of K at the second position, based on the observed specificity of NK, it was predicted that HK, IK, SK, and YK specify G.

[0264] These novel RVDs are tested and compared to existing RVDs for function and specificity in quantitative TAL effector and TALEN activity assays as described in Examples 2 and 11.

TABLE 71

RVDs grouped and ordered by their second residue

1 st residue	2 nd residue	Nucleotide
N	*	C or T
H	*	*
H	A	C
N	A	G
H	D	C
N	D	C
H	G	T
I	G	T
N	G	T
Y	G	T
N	I	A
H	I	C
N	K	G
H	N	G
S	N	G or A
N	N	G or A ¹
N	S	A or C or G ¹

¹An asterisk denotes a gap.

RVD groups with like specificities are boxed in thick lines.

Example 14

Custom TALENs Cleave Endogenous Targets in
Animal Cells and Introduce Mutations by Imprecise
Non-Homologous End Joining

[0265] To test whether TALENs could be used for targeted mutagenesis in animal cells, first, expression of TAL effectors AvrBs3, PthXo1, and Tal1c was tested in human embryonic kidney (HEK) 293T cells. The stop codon was removed from the AvrBs3, PthXo1, and Tal1c encoding genes and the genes were subcloned into mammalian expression vector pcDNA3.2/V5-DEST (Invitrogen, Carlsbad, Calif.) in frame with the downstream sequence in that vector that encodes the V5 epitope for protein immunodetection. pcDNA3.2/V5-DEST places the TAL effector gene under the control of the constitutive human cytomegalovirus (CMV) promoter. HEK 293T cells were transfected using Lipofectamine 2000 (Invitrogen) with the resulting plasmids individually, and after 24 hours, total proteins were isolated from each transfected batch of cells and subjected to polyacrylamide gel electrophoresis, western blotting and immunolabeling using a mouse anti-V5 antibody. The labeled proteins were detected with a goat anti-mouse antibody-horse radish peroxidase conjugate using the SuperSignal Weat Pico Chemiluminescent kit (ThermoScientific, Inc.). Equivalent loading was confirmed by immunolabeling and detection of actin. Each TAL effector protein was detectably expressed with no apparent degradation (FIG. 39).

[0266] Next, a pair of TALENs were designed as described in Example 9 to target a sequence in the endogenous human HPRT gene, and named HPRT-3254-17 and HPRT-3286-20r (FIG. 40A and FIG. 40B). Plasmids pTALEN141 encoding HPRT-3254-17 and plasmid pTALEN142 encoding HPRT-3286-20r were constructed using the Golden Gate cloning-based method and reagents described in Example 8. The

TALEN genes were then subcloned into the mammalian expression vector pcDNA3.1(−) (Invitrogen, Inc.), which places them under control of the constitutive CMV promoter, yielding plasmids pTALEN141M and pTALEN 142M. HEK 293T cells were then transfected with both pTALEN141M and pTALEN142M together and separately with pcDNA3.1(−) as a negative control. After 72 hours, genomic DNA was isolated and digested with restriction endonuclease Bpu10I. A Bpu10I site exists within the spacer that separates the HPRT-3254-17 and HPRT-3286-20r binding sites in HPRT (FIG. 41A). Following Bpu10I digestion, PCR was used to amplify a 244 bp fragment spanning the TALEN-targeted site from both the TALEN-treated and the control samples. The expected fragment was amplified from both samples, indicating that Bpu10I digestion of the genomic DNA had been incomplete. Subsequent digestion of the PCR products with Bpu10I, however, resulted in complete cleavage of the product amplified from the control sample, but incomplete cleavage of the product from the TALEN treated sample (FIG. 41B). The presence of cleavage-resistant PCR product in the TALEN-treated sample provides evidence that the endogenous Bpu10I site was mutated in vivo as a result of imperfect repair by non-homologous end joining of a TALEN-mediated double strand break at the intended target in HPRT. Thus, TALENs can be used for targeted mutagenesis in mammalian cells.

Other Embodiments

[0267] It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 107

<210> SEQ ID NO 1

<211> LENGTH: 34

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 1

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
20 25 30

His Gly

<210> SEQ ID NO 2

<211> LENGTH: 17

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 2

agaagaagag acccata

-continued

<210> SEQ ID NO 3
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 3

atataaacct aaccatcc 18

<210> SEQ ID NO 4
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 4

atataaacct gacccttt 18

<210> SEQ ID NO 5
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 5

atataaacct ctct 14

<210> SEQ ID NO 6
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 6

atataaacct aacca 15

<210> SEQ ID NO 7
<211> LENGTH: 14
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 7

ataaacctaa ccat 14

<210> SEQ ID NO 8
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 8

gcattctccc ctactgtaca ccac 24

<210> SEQ ID NO 9
<211> LENGTH: 23

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 9

ataaaaggcc ctcaccaacc cat                                23

<210> SEQ ID NO 10
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 10

ataatcccca aatccctcc tc                                  22

<210> SEQ ID NO 11
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 11

ccccctcget tccctt                                       16

<210> SEQ ID NO 12
<211> LENGTH: 1164
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas euvesicatoria

<400> SEQUENCE: 12

Met Asp Pro Ile Arg Ser Arg Thr Pro Ser Pro Ala Arg Glu Leu Leu
1      5      10      15

Pro Gly Pro Gln Pro Asp Gly Val Gln Pro Thr Ala Asp Arg Gly Val
20     25     30

Ser Pro Pro Ala Gly Gly Pro Leu Asp Gly Leu Pro Ala Arg Arg Thr
35     40     45

Met Ser Arg Thr Arg Leu Pro Ser Pro Pro Ala Pro Ser Pro Ala Phe
50     55     60

Ser Ala Gly Ser Phe Ser Asp Leu Leu Arg Gln Phe Asp Pro Ser Leu
65     70     75     80

Phe Asn Thr Ser Leu Phe Asp Ser Leu Pro Pro Phe Gly Ala His His
85     90     95

Thr Glu Ala Ala Thr Gly Glu Trp Asp Glu Val Gln Ser Gly Leu Arg
100    105    110

Ala Ala Asp Ala Pro Pro Pro Thr Met Arg Val Ala Val Thr Ala Ala
115    120    125

Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln Pro
130    135    140

Ser Asp Ala Ser Pro Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr
145    150    155    160

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
165    170    175

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
180    185    190

```

-continued

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val	195	200	205
Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala	210	215	220
Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala	225	230	235
Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp	245	250	255
Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val	260	265	270
Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn	275	280	285
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	290	295	300
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala	305	310	315
His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly	325	330	335
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	340	345	350
Gln Ala His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn	355	360	365
Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	370	375	380
Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala	385	390	395
Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	405	410	415
Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala	420	425	430
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Ala	435	440	445
Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val	450	455	460
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val	465	470	475
Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu	485	490	495
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu	500	505	510
Thr Val Gln Ala Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr	515	520	525
Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	530	535	540
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly	545	550	555
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	565	570	575
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala	580	585	590

-continued

His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly	595	600	605
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	610	615	620
Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn	625	630	640
Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val	645	650	655
Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala	660	665	670
Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	675	680	685
Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala	690	695	700
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	705	710	715
Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val	725	730	735
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val	740	745	750
Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu	755	760	765
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu	770	775	780
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr	785	790	795
Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Arg Pro Ala	805	810	815
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly	820	825	830
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	835	840	845
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala	850	855	860
His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly	865	870	875
Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp	885	890	895
Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys	900	905	910
Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His	915	920	925
Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr	930	935	940
Ser His Arg Val Ala Asp His Ala Gln Val Val Arg Val Leu Gly Phe	945	950	955
Phe Gln Cys His Ser His Pro Ala Gln Ala Phe Asp Asp Ala Met Thr	965	970	975
Gln Phe Gly Met Ser Arg His Gly Leu Leu Gln Leu Phe Arg Arg Val	980	985	990
Gly Val Thr Glu Leu Glu Ala Arg Ser Gly Thr Leu Pro Pro Ala Ser			

-continued

995	1000	1005
Gln Arg Trp Asp Arg Ile Leu Gln Ala Ser Gly Met Lys Arg Ala Lys		
1010	1015	1020
Pro Ser Pro Thr Ser Thr Gln Thr Pro Asp Gln Ala Ser Leu His Ala		
1025	1030	1035 1040
Phe Ala Asp Ser Leu Glu Arg Asp Leu Asp Ala Pro Ser Pro Met His		
	1045	1050 1055
Glu Gly Asp Gln Thr Arg Ala Ser Ser Arg Lys Arg Ser Arg Ser Asp		
	1060	1065 1070
Arg Ala Val Thr Gly Pro Ser Ala Gln Gln Ser Phe Glu Val Arg Val		
	1075	1080 1085
Pro Glu Gln Arg Asp Ala Leu His Leu Pro Leu Ser Trp Arg Val Lys		
	1090	1095 1100
Arg Pro Arg Thr Ser Ile Gly Gly Gly Leu Pro Asp Pro Gly Thr Pro		
1105	1110	1115 1120
Thr Ala Ala Asp Leu Ala Ala Ser Ser Thr Val Met Arg Glu Gln Asp		
	1125	1130 1135
Glu Asp Pro Phe Ala Gly Ala Ala Asp Asp Phe Pro Ala Phe Asn Glu		
	1140	1145 1150
Glu Glu Leu Ala Trp Leu Met Glu Leu Leu Pro Gln		
	1155	1160

<210> SEQ ID NO 13

<211> LENGTH: 4366

<212> TYPE: DNA

<213> ORGANISM: Xanthomonas euvesicatoria

<400> SEQUENCE: 13

```

gaattcaagg tgtcaaaaag cgataggcgg aattatagat gtacttgat gaacttatca    60
acgccagttt agtgaacggg ttcgacaaag cgaaaccaac acccaggcgc gaaagccttg    120
cgccgcaatg ctttcggcca atgtgaccca gggcattgac cgaaacggcg taggaatttc    180
ggaacacgac ggtaggggaa tgctctaccg cccggctacg caaaagcact ctgcctgcc    240
agacgcgcca ctgcgtggaa ttggcgtta tgtccgtgg cgccctcgcc gccgtagtgc    300
ttgcagcgca gccttgaatg atcgaaatc aaacatcact gtttgatagg tcgatcatga    360
catcgcccat ttcgagggtc ggcagggtt cgtgtaaaaa acagccaaaa gtgagctaac    420
tcgctgtcag cacagaaatt ttccacaacc ttctgccgat cctccatgcg ggtccgtgat    480
cgcttcctg tctgcgcctc accctggctg tcgagggttg ccaggatcac ccgaagtgtg    540
gtactgccat gcggcctcgg aagctatgta ggaaccacag accgctagtc tggaggcgac    600
catgtaaaga ggtatgcctg atggatccca ttcgttcgag cacaccaagt cctgcccgcg    660
agcttctgcc cggaccccaa cccgatgggg ttcagccgac tgcagatcgt ggggtgtctc    720
cgctgcggg cgccccctg gatggcttgc cgcctcgagg gacgatgtcc cggaaccggc    780
tgccatctcc ccctgcccc tcacctgcgt tctcggcggg cagcttcagt gacctgttac    840
gtcagttcga tccgtcactt ttaatacat cgctttttga ttcattgcct ccttcggcg    900
ctcaccatac agaggctgcc acaggcgagt gggatgaggt gcaatcgggt ctgcgggcag    960
ccgacgcccc cccaccacc atgcgcgtgg ctgtcactgc cgcgcggcgg ccgcgcgcca   1020
agcggcgccg gcgacgagct gctgcgcaac cctccgacgc ttcgcgggcc gcgcaggtgg   1080

```

-continued

atctacgcac gctcggctac agccagcagc aacaggagaa gatcaaaccg aaggttcgtt	1140
cgacagtggc gcagcaccac gaggcactgg tcggccatgg gtttacacac gcgcacatcg	1200
ttgcgctcag ccaacacccg gcagcgtagg ggaccgtcgc tgtcaagtat caggacatga	1260
tcgcagcgtt gccagaggcg acacacgaag cgatcgttgg cgtcggcaaa cagtgggccg	1320
gcgcacgcgc tctggaggcc ttgctcacgg tggcgggaga gttgagaggt ccaccgttac	1380
agttggacac aggccaaact ctcaagattg caaaacgtgg cggcgtgacc gcagtggagg	1440
cagtgcacgc atggcgcaat gcaactgacgg gtgccccct gaacctgacc ccggagcagg	1500
tgggtggcat cgccagccac gatggcggca agcaggcgct ggagacggtg cagcggctgt	1560
tgcgggtgct gtgccaggcc catggcctga ccccgagca ggtgggtggc atcgccagca	1620
atggcggtag caagcaggcg ctggagacgg tgcagcggct gttgccggtg ctgtgccagg	1680
cccatggcct gaccccgag caggtggtag ccatcgccag caatagcggg ggcaagcagg	1740
cgtggagac ggtgcagcgg ctggtgccgg tgcgtgtcca ggcccatggc ctgaccccg	1800
agcaggtggt ggccatcgcc agcaatggcg gtggcaagca ggcgctggag acggtgcagc	1860
ggctgttgcc ggtgctgtgc caggcccatg gcctgacccc ggagcaggtg gtggccatcg	1920
ccagcaatat tgggtgcaag caggcgtgag agcgggtgca ggcgctgttg ccggtgctgt	1980
gccaggccca tggcctgacc ccggagcagg tgggtggcat cgccagcaat attggtggca	2040
agcaggcgtg ggagacggtg caggcgtgtg tgcgggtgct gtgccaggcc catggcctga	2100
ccccggagca ggtgggtggc atcgccagca atattggtgg caagcaggcg ctggagacgg	2160
tgcagcgtg gttgccggtg ctgtgccagg cccatggcct gaccccgag caggtggtag	2220
ccatcgccag ccacgatggc ggcaagcagg cgctggagac ggtgcagcgg ctgttgccgg	2280
tgcgtgtcca ggcccatggc ctgaccccg agcagggtgt ggccatcgcc agccacgatg	2340
gcggcaagca ggcgctggag acggtgcagc ggctgttgcc ggtgctgtgc caggcccatg	2400
gcctgacccc gcagcaggtg gtggccatcg ccagcaatgg cggtaggcaag caggcgtg	2460
agacgggtgca gcggtgttg ccggtgctgt gccaggccca tggcctgacc ccggagcagg	2520
tgggtggcat cgccagcaat agcggtagga agcaggcgtg ggagacggtg caggcgtgt	2580
tgcgggtgct gtgccaggcc catggcctga ccccgagca ggtgggtggc atcgccagca	2640
atagcggtag caagcaggcg ctggagacgg tgcagcggct gttgccggtg ctgtgccagg	2700
cccatggcct gaccccgag caggtggtag ccatcgccag ccacgatggc ggcaagcagg	2760
cgtggagac ggtgcagcgg ctggtgccgg tgcgtgtcca ggcccatggc ctgaccccg	2820
agcaggtggt ggccatcgcc agccacgatg gcggcaagca ggcgctggag acggtgcagc	2880
ggctgttgcc ggtgctgtgc caggcccatg gcctgacccc ggagcaggtg gtggccatcg	2940
ccagccacga tggcggaag caggcgtgag agcgggtgca ggcgctgttg ccggtgctgt	3000
gccaggccca tggcctgacc ccgcagcagg tgggtggcat cgccagcaat ggcgcgaggca	3060
ggcgcggtg ggagacggtg cagcggctgt tgcgggtgct gtgccaggcc catggcctga	3120
ccccggagca ggtgggtggc atcgccagcc acgatggcgg caagcaggcg ctggagacgg	3180
tgcagcggct gttgccggtg ctgtgccagg cccatggcct gaccccgag caggtggtag	3240
ccatcgccag caatggcggc gcagggcgg cgctggagag cattgttgcc cagttatctc	3300
gcctgatcc ggcttggcc gcgttgacca acgaccacct cgtcgccctg gcctgcctcg	3360

-continued

```

gcggacgtcc tgcgctggat gcagtga aaa agggattgcc gcacgcgccg gccttgatca 3420
aaagaaccaa tcgccgtatt cccgaacgca catcccatcg cgttgccgac cagcgcaag 3480
tggttcgcgt gctgggtttt ttccagtgcc actccacccc agcgcaagca tttgatgacg 3540
ccatgacgca gttcgggatg agcaggcacg ggttggtaca gctctttcgc agagtgggcg 3600
tcaccgaact cgaagcccgc agtggaacgc tccccccagc ctgcgagcgt tgggaccgta 3660
tcctccaggc atcagggatg aaaagggcc aaccgtcccc tacttcaact caaacgccgg 3720
atcaggcgtc tttgcatgca ttgcgcgatt cgctggagcg tgaccttgat gcgcctagcc 3780
caatgcacga gggagatcag acgcgggcaa gcagccgtaa acggtcccga tcggatcgtg 3840
ctgtcacccg tccctccgca cagcaatcgt tcgaggtgcg cgttcccga cagcgcgatg 3900
cgctgcattt gcccctcagt tggagggtaa aacgcccgcg taccagtatc gggggcgggc 3960
tcccgatcc tggtagcccc acggctgccg acctggcagc gtccagcacc gtgatgcggg 4020
aacaagatga ggaccccttc gcaggggcag cggatgattt cccggcattc aacgaagagg 4080
agctcgcatg gttgatggag ctattgcctc agtgaggctc agtcggtgac tacctgagcg 4140
tcggcaggga ttggtgtaag taacctttac tgacagcgag ttagcccaact tttggctgtt 4200
ttttacacaa atccctgcct cccctctggt tgcaccacac ccgtacacca agcgcggcgg 4260
cgaagcaggc accgagtggg tccgctgcgg tgttgcgttc cctaaccagg gcggtggcta 4320
tacgctcaag ctgcgcaccg tcccgggtggc gatcgacgac gaaatg 4366

```

```

<210> SEQ ID NO 14
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

```

```

<400> SEQUENCE: 14

```

```

tatataaacc taaccatcct cacaacttca agttatcgga tggttagggt tatata 56

```

```

<210> SEQ ID NO 15
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: target sequence

```

```

<400> SEQUENCE: 15

```

```

tatataaacc taaccatccg ataacttgaa gttgtgagga tggttagggt tatata 56

```

```

<210> SEQ ID NO 16
<211> LENGTH: 1373
<212> TYPE: PRT
<213> ORGANISM: Xanthomonas euvesicatoria

```

```

<400> SEQUENCE: 16

```

```

Met Asp Pro Ile Arg Ser Arg Thr Pro Ser Pro Ala Arg Glu Leu Leu
1           5           10          15

```

```

Pro Gly Pro Gln Pro Asp Arg Val Gln Pro Thr Ala Asp Arg Gly Gly
20          25          30

```

```

Ala Pro Pro Ala Gly Gly Pro Leu Asp Gly Leu Pro Ala Arg Arg Thr
35          40          45

```

```

Met Ser Arg Thr Arg Leu Pro Ser Pro Pro Ala Pro Ser Pro Ala Phe

```

-continued

50	55	60
Ser Ala Gly Ser Phe	Ser Asp Leu Leu Arg	Gln Phe Asp Pro Ser Leu
65	70	75 80
Leu Asp Thr Ser Leu	Leu Asp Ser Met Pro	Ala Val Gly Thr Pro His
	85	90 95
Thr Ala Ala Ala Pro	Ala Glu Cys Asp Glu	Val Gln Ser Gly Leu Arg
	100	105 110
Ala Ala Asp Asp Pro	Pro Pro Thr Val Arg	Val Ala Val Thr Ala Ala
	115	120 125
Arg Pro Pro Arg Ala	Lys Pro Ala Pro Arg	Arg Arg Ala Ala Gln Pro
	130	135 140
Ser Asp Ala Ser Pro	Ala Ala Gln Val Asp	Leu Arg Thr Leu Gly Tyr
145	150	155 160
Ser Gln Gln Gln Gln	Glu Lys Ile Lys Pro	Lys Val Gly Ser Thr Val
	165	170 175
Ala Gln His His Glu	Ala Leu Val Gly His	Gly Phe Thr His Ala His
	180	185 190
Ile Val Ala Leu Ser	Arg His Pro Ala Ala	Leu Gly Thr Val Ala Val
	195	200 205
Lys Tyr Gln Asp Met	Ile Ala Ala Leu Pro	Glu Ala Thr His Glu Asp
	210	215 220
Ile Val Gly Val Gly	Lys Gln Trp Ser Gly	Ala Arg Ala Leu Glu Ala
225	230	235 240
Leu Leu Thr Val Ala	Gly Glu Leu Arg Gly	Pro Pro Leu Gln Leu Asp
	245	250 255
Thr Gly Gln Leu Val	Lys Ile Ala Lys Arg	Gly Gly Val Thr Ala Val
	260	265 270
Glu Ala Val His Ala	Ser Arg Asn Ala Leu	Thr Gly Ala Pro Leu Asn
	275	280 285
Leu Thr Pro Ala Gln	Val Val Ala Ile Ala	Ser Asn Asn Gly Gly Lys
	290	295 300
Gln Ala Leu Glu Thr	Val Gln Arg Leu Leu	Pro Val Leu Cys Gln Ala
305	310	315 320
His Gly Leu Thr Pro	Ala Gln Val Val Ala	Ile Ala Ser His Asp Gly
	325	330 335
Gly Lys Gln Ala Leu	Glu Thr Met Gln Arg	Leu Leu Pro Val Leu Cys
	340	345 350
Gln Ala His Gly Leu	Pro Pro Asp Gln Val	Val Ala Ile Ala Ser Asn
	355	360 365
Ile Gly Gly Lys Gln	Ala Leu Glu Thr Val	Gln Arg Leu Leu Pro Val
	370	375 380
Leu Cys Gln Ala His	Gly Leu Thr Pro Asp	Gln Val Val Ala Ile Ala
385	390	395 400
Ser His Gly Gly Gly	Lys Gln Ala Leu Glu	Thr Val Gln Arg Leu Leu
	405	410 415
Pro Val Leu Cys Gln	Ala His Gly Leu Thr	Pro Asp Gln Val Val Ala
	420	425 430
Ile Ala Ser His Asp	Gly Gly Lys Gln Ala	Leu Glu Thr Val Gln Arg
	435	440 445
Leu Leu Pro Val Leu	Cys Gln Ala His Gly	Leu Thr Pro Asp Gln Val
	450	455 460

-continued

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

-continued

Gly 865	Leu	Thr	Pro	Ala	Gln 870	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly 880	Gly
Lys	Gln	Ala	Leu	Glu 885	Thr	Val	Gln	Arg	Leu 890	Leu	Pro	Val	Leu	Cys	Gln 895
Asp	His	Gly	Leu 900	Thr	Leu	Ala	Gln	Val 905	Val	Ala	Ile	Ala	Ser	Asn	Ile 910
Gly	Gly	Lys 915	Gln	Ala	Leu	Glu	Thr	Val 920	Gln	Arg	Leu	Leu	Pro	Val	Leu 925
Cys	Gln	Ala	His	Gly	Leu 930	Thr	Gln	Asp 935	Gln	Val	Val	Ala	Ile	Ala	Ser 940
Asn	Ile	Gly	Gly	Lys	Gln 945	Ala	Leu	Glu	Thr	Val 950	Gln	Arg	Leu	Leu	Pro 955
Val	Leu	Cys	Gln	Asp 960	His	Gly	Leu	Thr 965	Pro	Asp	Gln	Val	Val	Ala	Ile 970
Ala	Ser	Asn	Ile	Gly 975	Gly	Lys	Gln	Ala 980	Leu	Glu	Thr	Val	Gln	Arg	Leu 985
Leu	Pro	Val	Leu	Cys 990	Gln	Asp	His	Gly 995	Leu	Thr	Leu	Asp	Gln	Val	Val 1000
Ala	Ile	Ala	Ser	Asn	Gly 1005	Gly	Lys	Gln 1010	Ala	Leu	Glu	Thr	Val	Gln	Arg 1015
Leu	Leu	Pro	Val	Leu	Cys 1020	Gln	Asp	His 1025	Gly	Leu	Thr	Pro	Asp	Gln	Val 1030
Val	Ala	Ile	Ala	Ser	Asn 1035	Ser	Gly	Gly 1040	Lys	Gln	Ala	Leu	Glu	Thr	Val 1045
Gln	Arg	Leu	Leu	Pro	Val 1050	Leu	Cys	Gln 1055	Asp	His	Gly	Leu	Thr	Pro	Asn 1060
Gln	Val	Val	Ala	Ile	Ala 1065	Ser	Asn	Gly 1070	Gly	Lys	Gln	Ala	Leu	Glu	Ser 1075
Ile	Val	Ala	Gln	Leu	Ser 1080	Arg	Pro	Asp 1085	Pro	Ala	Leu	Ala	Ala	Leu	Thr 1090
Asn	Asp	His	Leu	Val	Ala 1095	Leu	Ala	Cys 1100	Leu	Gly	Gly	Arg	Pro	Ala	Met 1105
Asp	Ala	Val	Lys	Lys	Gly 1110	Leu	Pro	His 1115	Ala	Pro	Glu	Leu	Ile	Arg	Arg 1120
Val	Asn	Arg	Arg	Ile	Gly 1125	Glu	Arg	Thr 1130	Ser	His	Arg	Val	Ala	Asp	Tyr 1135
Ala	Gln	Val	Val	Arg	Val 1140	Leu	Glu	Phe 1145	Phe	Gln	Cys	His	Ser	His	Pro 1150
Ala	Tyr	Ala	Phe	Asp	Glu 1155	Ala	Met	Thr 1160	Gln	Phe	Gly	Met	Ser	Arg	Asn 1165
Gly	Leu	Val	Gln	Leu	Phe 1170	Arg	Arg	Val 1175	Gly	Val	Thr	Glu	Leu	Glu	Ala 1180
Arg	Gly	Gly	Thr	Leu	Pro 1185	Pro	Ala	Ser 1190	Gln	Arg	Trp	Asp	Arg	Ile	Leu 1195
Gln	Ala	Ser	Gly	Met	Lys 1200	Arg	Ala	Lys 1205	Pro	Ser	Pro	Thr	Ser	Ala	Gln 1210
Thr	Pro	Asp	Gln	Ala	Ser 1215	Leu	His	Ala 1220	Phe	Ala	Asp	Ser	Leu	Glu	Arg 1225
Asp	Leu	Asp	Ala	Pro	Ser 1230	Pro	Met	His 1235	Glu	Gly	Asp	Gln	Thr	Gly	Ala 1240
Ser	Ser	Arg	Lys	Arg	Ser 1245	Arg	Ser	Asp 1250	Arg	Ala	Val	Thr	Gly	Pro	Ser 1255

-continued

1265	1270	1275	1280
Ala Gln His Ser Phe Glu Val Arg Val Pro Glu Gln Arg Asp Ala Leu			
	1285	1290	1295
His Leu Pro Leu Ser Trp Arg Val Lys Arg Pro Arg Thr Arg Ile Gly			
	1300	1305	1310
Gly Gly Leu Pro Asp Pro Gly Thr Pro Ile Ala Ala Asp Leu Ala Ala			
	1315	1320	1325
Ser Ser Thr Val Met Trp Glu Gln Asp Ala Ala Pro Phe Ala Gly Ala			
	1330	1335	1340
Ala Asp Asp Phe Pro Ala Phe Asn Glu Glu Glu Leu Ala Trp Leu Met			
	1345	1350	1355
			1360
Glu Leu Leu Pro Gln Ser Gly Ser Val Gly Gly Thr Ile			
	1365	1370	

<210> SEQ ID NO 17
 <211> LENGTH: 102
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 17

```

ctgaccccg cagaggtggt ggccatcgcc agcmaygayg gcggcaagca ggcgctggag      60
acgggtgcagc ggctgtgtgcc ggtgctgtgc caggaccatg gc                        102
  
```

<210> SEQ ID NO 18
 <211> LENGTH: 34
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 18

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys			
1	5	10	15
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp			
	20	25	30

His Gly

<210> SEQ ID NO 19
 <211> LENGTH: 63
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 19

```

ctgaccccg cagaggtggt ggccatcgcc agcmaygayg gcggcaagca ggcgctcgag      60
agc                                                                63
  
```

<210> SEQ ID NO 20
 <211> LENGTH: 21
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 20

-continued

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Ser
20

<210> SEQ ID NO 21
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 21

gcgctggaga gc 12

<210> SEQ ID NO 22
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 22

Ala Leu Glu Ser
1

<210> SEQ ID NO 23
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 23

gcgctcgagt cc 12

<210> SEQ ID NO 24
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 24

tcgagacggt gcagcggctg ttgccgtgc tgtgccagga ccatggcctg accccggacc 60

aagtgggtggc catcgccagc aacattggcg gcaagcaagc gc 102

<210> SEQ ID NO 25
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 25

tcgagcgctt gcttgccgcc aatgttgctg gcgatggcca ccacttggtc cggggtcagg 60

ccatggtcct ggcacagcac cggcaacagc cgctgcaccg tc 102

<210> SEQ ID NO 26
<211> LENGTH: 35
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 26

-continued

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
 1 5 10 15
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 20 25 30
 Ala Leu Glu
 35

<210> SEQ ID NO 27
 <211> LENGTH: 102
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 27

tcgaaacggt gcagcggctg ttgccgtgc tgtgccagga ccatggcctg accccggacc 60
 aagtgtggc tatcgccagc aacattggcg gcaagcaagc gc 102

<210> SEQ ID NO 28
 <211> LENGTH: 102
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 28

tcgagcgctt gcttgccgcc aatgttgctg gcgatagcca ccacttggtc cggggtcagg 60
 ccatggtcct ggcacagcac cggcaacagc cgctgcaccg tt 102

<210> SEQ ID NO 29
 <211> LENGTH: 18
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 29

atcaagattc tcttcact 18

<210> SEQ ID NO 30
 <211> LENGTH: 15
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: target sequence

<400> SEQUENCE: 30

cccagaagta aacat 15

<210> SEQ ID NO 31
 <211> LENGTH: 598
 <212> TYPE: PRT
 <213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 31

Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 1 5 10 15
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
 20 25 30

-continued

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

-continued

435	440	445
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val		
450	455	460
Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu		
465	470	475
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu		
	485	490
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr		
	500	505
Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Arg Pro Ala		
	515	520
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly		
	530	535
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys		
	545	550
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala		
	565	570
His Gly Leu Thr Pro Gln Gln Val Val Ala Ile Ala Ser Asn Gly Gly		
	580	585
Gly Arg Pro Ala Leu Glu		
	595	

<210> SEQ ID NO 32

<211> LENGTH: 4122

<212> TYPE: DNA

<213> ORGANISM: Xanthomonas oryzae

<400> SEQUENCE: 32

```

atggatccca ttcgttcgcg caccgccaagt cctgcccgcg agcttctgcc cggaccccaa    60
ccgatagggg ttcagccgac tgcagatcgg gggggggctc cgctctgtgg cggccccctg    120
gatggcttgc ccgctcggcg gacgatgtcc cggacccggc tgccatctcc cctgcgccc    180
tcgctcgctg tctcggcggg cagcttcagc gatctgtccc gtcagttega tccgtcgctt    240
cttgatacat cgcttcttga ttcgatgcct gccgtcggca cgcgcatac agcggctgcc    300
ccagcagagt gcatgagagt gcaatcgggt ctgctgtcag ccgatgaccc gccaccacc    360
gtcgtgtgct ctgtcactgc cgcgcggcgg ccgcgcgcca agccggcccc gcgacggcgt    420
gcggcgcaac cctccgacgc ttcgcgggcc gcgcaggtgg atctacgcac gctcggttac    480
agtcagcagc agcaagagaa gatcaaaccg aaggtggggt cgacagtggc gcagcaccac    540
gaggcactgg tggggccatgg gtttacacac gcgcacatcg ttgcgctcag ccgacacccg    600
gcagcgttag ggaccgtcgc tgtcaagtat caggacatga tcgcggcggt accagaggcg    660
acacacgaag acatcgttgg tgtcggcaaa cagtgggtccg gcgcacgcgc cctggaggcc    720
ttgtcacagg tggcgggaga gttgagaggt ccaccgttac agttggacac aggccaactt    780
gtcaagattg caaaacgtgg cggcgtgacc gcagtggagg cagtgcacgc atcgcgcaat    840
gcactgacgg gtgccccctt gaacctgacc ccggcacagg tgggtggccat cggcagcaat    900
aacggtgcca agcaggcgct ggagacgggt cagcggctgt tgccgggtgt gtgccaggcc    960
catggcctga ccccgcgcca ggtggtggcc atcgccagcc acgatggcgg caagcaggca   1020
ctggagacga tgcagcgggt gttgcgggtg ctgtgccagg cccatggcct gccccgggac   1080

```

-continued

caggtggtgg	ccatcgccag	caatattggc	ggcaagcagg	cgctggagac	ggtgcagcgg	1140
ctgttgccgg	tgctctgcca	ggcccatggc	ctgaccccg	accaggtggt	ggccatcgcc	1200
agccatggcg	gcggaagca	ggcgctggag	acggtgcagc	ggctgttgcc	ggtgctctgc	1260
caggcccatg	gcctgacccc	ggaccaggtg	gtggccatcg	ccagccacga	tggcggcaag	1320
caggcgctgg	agacggtgca	gcggctgttg	cgggtgctgt	gccaggccca	tggcctgacc	1380
ccggaccagg	tggtgcccat	cgccagcaat	ggcgccggca	agcaggcgct	ggagacggtg	1440
cagcggtgtg	tgccggtgct	gtgccaggcc	catggtctga	ccccggacca	ggtggtggcc	1500
atcgccagca	atggcgccaa	gcaggcgctg	gagacggtgc	agcggtctgt	gccggtgctg	1560
tggcaggccc	atggcctgac	cccgaccag	gtggtggcca	tgcgagcca	cgatggcggc	1620
aagcaggcgc	tggagacggt	gcagcggtg	ttgccggtgc	tgtgccagac	ccatggtctg	1680
accccgcg	aggtggtggc	catcgccagc	cacgatggcg	gcaagcaggc	gctggagacg	1740
gtgcagcagc	tgttgccggt	gctgtgccag	gcccattggc	tgaccccgga	ccaggtggtg	1800
gccatcgcca	gcaatattgg	cggcaagcag	gcgctagcga	cgggtgcagc	gctgttgccg	1860
gtgctgtgcc	aagccccatg	cctgaccccg	gaccaggtgg	tggccatcgc	cagcaatggc	1920
ggcggaagc	agggcggtga	gacggtgcag	cggctgttgc	cgggtgctgtg	ccaggcccat	1980
ggcctgaccc	cggaccaggt	ggtggccatc	gccagcaatg	gcggcgga	gcaggcgctg	2040
gagacggtgc	agcggtctgt	gccggtgctg	tgcaggccc	atggtctgac	ccaggtgcag	2100
gtggtggcca	tgcgagcaa	tattggcggc	aagcaggcgc	tggagacggt	gcagcggtg	2160
ttgccggtgc	tgtgccaggc	ccatggcctg	accccgcg	aggtggtggc	catcgccagc	2220
cacgatggcg	gcaagcaggc	gctggagacg	gtgcagcggc	tgttgccggt	gctgtgccag	2280
gcccattggc	tgaccccgga	ccaagtgggt	gccatcgcca	gcaatggcgg	cggcaagcag	2340
gcgctggaga	cgggtgcagc	gctgttgccg	gtgctgtgcc	aggcccatgg	cctgacccag	2400
gagcaggtgg	tggccatcgc	cagcaataac	ggcggaagc	aggcgctgga	gacggtgcag	2460
cggctgttgc	cgggtgctgtg	ccaggcccat	ggcctgaccc	cggaccaggt	ggtggccatc	2520
gccagcaatg	gcggcgga	gcaggcgctg	gagacggtgc	agcggtctgt	gccggtgctg	2580
tgcaggccc	atggtctgac	cccgcgag	gtggtggcca	tgcgagcaa	tattggcggc	2640
aagcaggcgc	tggagacggt	gcagcggtg	ttgccggtgc	tgtgccagga	ccatggcctg	2700
accctggcgc	aggtggtggc	catcgccagc	aatattggcg	gcaagcaggc	gctggagacg	2760
gtgcagcggc	tgttgccggt	gctgtgccag	gcacatggcc	tgaccagga	ccaggtggtg	2820
gccatcgcca	gcaatattgg	cggcaagcag	gcgctggaga	cgggtgcagc	gctgttgccg	2880
gtgctgtgcc	aggaccatgg	cctgaccccg	gaccaggtcg	tggccatcgc	cagcaatatt	2940
ggcggaagc	agggcggtga	gacggtgcag	cggctgttgc	cgggtgctgtg	ccaggaccat	3000
ggcctgaccc	tggaccaggt	ggtggccatc	gccagcaatg	gcggcaagca	ggcgctggag	3060
acggtgcagc	ggctgttgcc	ggtgctgtgc	caggaccatg	gactgacccc	ggaccaggtc	3120
gtggccatcg	ccagcaatag	tggcggaag	caggcgctgg	agacggtgca	gcggctgttg	3180
ccggtgctgt	gccaggacca	tggcctgacc	ccgaaccagg	tgggtggccat	cgccagcaat	3240
ggcggaagc	agggcggtga	gagcattgtt	gccagttat	ctcgccctga	tccggcggtg	3300
gcccggttga	ccaacgacca	cctcgctgcc	ttggcctgcc	tccggcgacg	tcccgccatg	3360

-continued

```

gatgcagtga aaaagggatt gccgcacgcg ccggaattga tcagaagagt caatcgccgt 3420
attggcgaac gcacgtccca tcgcgttgcc gactacgcgc aagtggttcg cgtgctggag 3480
tttttcagt gccactccca ccacgcgtac gcatttgatg aggccatgac gcagttcggg 3540
atgagcagga acgggttggt acagctcttt cgcagagtgg gcgtcaccga actcgaagcc 3600
cgcggtggaa cgctccccc agcctcgcag cggtgggacc gtatcctcca ggcatcaggg 3660
atgaaaaggg ccaaaccgtc ccctacttca gctcaaacac cggatcaggc gtctttgcat 3720
gcattcgccg attcgctgga gcgtgacctt gatgcgccta gccaatgca cgagggagat 3780
cagacagggg caagcagccg taaacggtcc cgatcggatc gtgctgtcac cggcccctcc 3840
gcacagcact ctttcagagt gcgcgttccc gaacagcgcg atcgctgca tttgcccctc 3900
agctggaggg taaaacgccc gcgtaccagg atcgggggcg gcctcccgga tcctggtacg 3960
cccatcgctg ccgacctggc agcgtccagc accgtgatgt gggaaacaaga tgcggccccc 4020
ttcgagggg cagcggatga ttcccgga ttcaacgaag aggagctgc atggttgatg 4080
gagctattgc ctcagtcagg ctcagtcgga gggacgatct ga 4122

```

```

<210> SEQ ID NO 33
<211> LENGTH: 1341
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

```

```

<400> SEQUENCE: 33

```

```

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1           5           10          15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Ser Arg Thr
20          25          30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Gly Val
35          40          45

Gln Pro Thr Ala Asp Arg Gly Val Ser Pro Pro Ala Gly Gly Pro Leu
50          55          60

Asp Gly Leu Pro Ala Arg Arg Thr Met Ser Arg Thr Arg Leu Pro Ser
65          70          75          80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85          90          95

Leu Arg Gln Phe Asp Pro Ser Leu Phe Asn Thr Ser Leu Phe Asp Ser
100         105         110

Leu Pro Pro Phe Gly Ala His His Thr Glu Ala Ala Thr Gly Glu Trp
115         120         125

Asp Glu Val Gln Ser Gly Leu Arg Ala Ala Asp Ala Pro Pro Pro Thr
130         135         140

Met Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
145         150         155         160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
165         170         175

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
180         185         190

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
195         200         205

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro

```

-continued

210	215	220
Ala Ala Leu Gly Thr Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala		
225	230	235 240
Leu Pro Glu Ala Thr His Glu Ala Ile Val Gly Val Gly Lys Gln Trp		
	245	250 255
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu		
	260	265 270
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala		
	275	280 285
Lys Arg Gly Gly Val Thr Ala Val Glu Ala Val His Ala Trp Arg Asn		
	290	295 300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Glu Gln Val Val Ala		
305	310	315 320
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
	325	330 335
Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Gln Gln Val		
	340	345 350
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val		
	355	360 365
Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Gln		
	370	375 380
Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu		
385	390	395 400
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr		
	405	410 415
Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala		
	420	425 430
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly		
	435	440 445
Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys		
	450	455 460
Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val Leu Cys Gln Ala		
465	470	475 480
His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly		
	485	490 495
Gly Lys Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val Leu Cys		
	500	505 510
Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn		
	515	520 525
Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Ala Leu Leu Pro Val		
	530	535 540
Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala		
545	550	555 560
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
	565	570 575
Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Glu Gln Val Val Ala		
	580	585 590
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
	595	600 605
Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Gln Gln Val		
	610	615 620

-continued

Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	625	630	635	640
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	645	650	655	
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	660	665	670	
Thr	Val	Gln	Ala	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	675	680	685	
Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	690	695	700	
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	705	710	715	720
Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	725	730	735	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	740	745	750	
His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	755	760	765	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	770	775	780	
Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser	His	785	790	795	800
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	805	810	815	
Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Gln	Gln	Val	Val	Ala	Ile	Ala	820	825	830	
Ser	Asn	Gly	Gly	Gly	Arg	Pro	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	835	840	845	
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Glu	Gln	Val	Val	Ala	850	855	860	
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	865	870	875	880
Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Gln	Gln	Val	885	890	895	
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Arg	Pro	Ala	Leu	Glu	Ser	Ile	900	905	910	
Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	915	920	925	
Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	930	935	940	
Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Ala	Leu	Ile	Lys	Arg	Thr	945	950	955	960
Asn	Arg	Arg	Ile	Pro	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	His	Ala	965	970	975	
Gln	Val	Val	Arg	Val	Leu	Gly	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	980	985	990	
Gln	Ala	Phe	Asp	Asp	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	995	1000	1005	
Leu	Leu	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Leu	Glu	Ala	Arg	1010	1015	1020	

-continued

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

<210> SEQ ID NO 34

<211> LENGTH: 1542

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 34

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20 25 30

-continued

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

-continued

435						440						445					
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys		
450						455						460					
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala		
465						470						475					
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly		
						485						490					
												495					
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys		
						500						505					
												510					
Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn		
						515						520					
												525					
Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu		
530						535						540					
Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser		
545						550						555					
												560					
His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro		
						565						570					
												575					
Val	Leu	Cys	Gln	Thr	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile		
						580						585					
												590					
Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Gln	Leu		
						595						600					
												605					
Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val		
610						615						620					
Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Ala	Thr	Val	Gln		
625						630						635					
												640					
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln		
						645						650					
												655					
Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr		
						660						665					
												670					
Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro		
						675						680					
												685					
Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu		
690						695						700					
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu		
705						710						715					
												720					
Thr	Gln	Val	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln		
						725						730					
												735					
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His		
						740						745					
												750					
Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly		
						755						760					
												765					
Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln		
770						775						780					
Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly		
785						790						795					
												800					
Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu		
						805						810					
												815					
Cys	Gln	Ala	His	Gly	Leu	Thr	Gln	Glu	Gln	Val	Val	Ala	Ile	Ala	Ser		
						820						825					
												830					
Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro		
						835						840					
												845					

-continued

Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile
850						855					860				
Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu
865					870					875					880
Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val
				885					890					895	
Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln
			900					905						910	
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Leu	Ala	Gln
			915				920						925		
Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr
	930						935					940			
Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Gln
945					950					955					960
Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu
				965						970					975
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu
			980						985					990	
Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln
			995				1000						1005		
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His
	1010						1015					1020			
Gly	Leu	Thr	Leu	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Lys
1025					1030					1035					1040
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
				1045						1050					1055
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly
		1060						1065						1070	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
		1075						1080					1085		
Gln	Asp	His	Gly	Leu	Thr	Pro	Asn	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn
	1090						1095					1100			
Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro
1105					1110					1115					1120
Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala
				1125					1130					1135	
Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro
		1140						1145						1150	
His	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg
	1155							1160					1165		
Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	Gln	Val	Val	Arg	Val	Leu	Glu
	1170						1175					1180			
Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	Tyr	Ala	Phe	Asp	Glu	Ala	Met
1185						1190				1195					1200
Thr	Gln	Phe	Gly	Met	Ser	Arg	Asn	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg
				1205					1210					1215	
Val	Gly	Val	Thr	Glu	Leu	Glu	Ala	Arg	Gly	Gly	Thr	Leu	Pro	Pro	Ala
			1220					1225					1230		
Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala
		1235					1240						1245		

-continued

Lys Pro Ser Pro Thr Ser Ala Gln Thr Pro Asp Gln Ala Ser Leu His
 1250 1255 1260
 Ala Phe Ala Asp Ser Leu Glu Arg Asp Leu Asp Ala Pro Ser Pro Met
 1265 1270 1275 1280
 His Glu Gly Asp Gln Thr Arg Ala Ser Ser Arg Lys Arg Ser Arg Ser
 1285 1290 1295
 Asp Arg Ala Val Thr Gly Pro Ser Ala Gln Gln Ala Val Glu Val Arg
 1300 1305 1310
 Val Pro Glu Gln Arg Asp Ala Leu His Leu Pro Leu Ser Trp Arg Val
 1315 1320 1325
 Lys Arg Pro Arg Thr Arg Ile Trp Gly Gly Leu Pro Asp Pro Ile Ser
 1330 1335 1340
 Arg Ser Gln Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu
 1345 1350 1355 1360
 Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu
 1365 1370 1375
 Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met
 1380 1385 1390
 Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly
 1395 1400 1405
 Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp
 1410 1415 1420
 Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu
 1425 1430 1435 1440
 Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn Gln
 1445 1450 1455
 Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro
 1460 1465 1470
 Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys
 1475 1480 1485
 Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys
 1490 1495 1500
 Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met
 1505 1510 1515 1520
 Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn
 1525 1530 1535
 Asn Gly Glu Ile Asn Phe
 1540

<210> SEQ ID NO 35
 <211> LENGTH: 1035
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 35

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
 1 5 10 15
 Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
 20 25 30
 Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
 35 40 45

-continued

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

-continued

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

-continued

```

Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu
865                870                875                880

Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly
                885                890                895

Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val
                900                905                910

Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser
                915                920                925

Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr
                930                935                940

Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp
945                950                955                960

Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val
                965                970                975

Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn
                980                985                990

His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu
                995                1000                1005

Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val
1010                1015                1020

Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
1025                1030                1035

```

```

<210> SEQ ID NO 36
<211> LENGTH: 1069
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

```

```

<400> SEQUENCE: 36

```

```

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1      5      10      15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20     25     30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35     40     45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
50     55     60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
65     70     75     80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85     90     95

Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
100    105    110

Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
115    120    125

Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
130    135    140

Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
145    150    155    160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
165    170    175

```

Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	
			180							185				190		
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	
			195				200							205		
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	
			210				215							220		
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	
			225				230				235				240	
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	
			245						250						255	
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	
			260						265						270	
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	
			275						280						285	
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	
			290			295			300							
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	
			305			310			315						320	
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
			325						330						335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
			340						345						350	
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
			355			360						365				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
			370			375			380							
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
			385			390			395						400	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			405						410						415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
			420						425						430	
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
			435			440						445				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
			450			455			460							
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
			465			470			475						480	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
			485						490						495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
			500						505						510	
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
			515			520						525				
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
			530			535						540				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
			545			550			555						560	
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
			565						57							

-continued

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

-continued

980				985				990							
Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu
	995						1000						1005		
Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg
	1010						1015						1020		
Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu
	1025					1030					1035				1040
Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu
			1045							1050					1055
Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe			
		1060								1065					
<210> SEQ ID NO 37															
<211> LENGTH: 1137															
<212> TYPE: PRT															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: synthetic peptide															
<400> SEQUENCE: 37															
Met	Ala	Ser	Ser	Pro	Pro	Lys	Lys	Lys	Arg	Lys	Val	Ser	Trp	Lys	Asp
1				5					10					15	
Ala	Ser	Gly	Trp	Ser	Arg	Met	His	Ala	Asp	Pro	Ile	Arg	Pro	Arg	Arg
		20					25						30		
Pro	Ser	Pro	Ala	Arg	Glu	Leu	Leu	Pro	Gly	Pro	Gln	Pro	Asp	Arg	Val
		35				40							45		
Gln	Pro	Thr	Ala	Asp	Arg	Gly	Val	Ser	Ala	Pro	Ala	Gly	Ser	Pro	Leu
	50					55					60				
Asp	Gly	Leu	Pro	Ala	Arg	Arg	Thr	Val	Ser	Arg	Thr	Arg	Leu	Pro	Ser
	65				70				75					80	
Pro	Pro	Ala	Pro	Ser	Pro	Ala	Phe	Ser	Ala	Gly	Ser	Phe	Ser	Asp	Leu
			85						90					95	
Leu	Arg	Pro	Phe	Asp	Pro	Ser	Leu	Leu	Asp	Thr	Ser	Leu	Leu	Asp	Ser
		100					105						110		
Met	Pro	Ala	Val	Gly	Thr	Pro	His	Thr	Ala	Ala	Ala	Pro	Ala	Glu	Trp
		115				120							125		
Asp	Glu	Ala	Gln	Ser	Ala	Leu	Arg	Ala	Ala	Asp	Asp	Pro	Pro	Pro	Thr
	130					135					140				
Val	Arg	Val	Ala	Val	Thr	Ala	Ala	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala
	145				150					155				160	
Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln
			165						170					175	
Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile
		180					185						190		
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val
		195					200						205		
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro
	210					215					220				
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala
	225				230					235				240	
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp
			245						250					255	
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu

-continued

260							265							270						
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala					
	275						280					285								
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn					
	290					295					300									
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala					
305					310					315					320					
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg					
			325					330						335						
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val					
		340						345					350							
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val					
	355					360						365								
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp					
	370					375					380									
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu					
385					390					395					400					
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr					
			405					410						415						
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala					
		420					425						430							
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly					
	435					440					445									
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys					
	450				455						460									
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp					
465					470					475					480					
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly					
			485					490						495						
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys					
		500					505						510							
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn					
	515					520						525								
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val					
	530				535					540										
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala					
545					550					555					560					
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu					
			565					570						575						
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala					
		580					585						590							
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg					
		595				600						605								
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val					
	610					615						620								
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val					
625					630					635					640					
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp					
			645						650					655						
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu					
		660					665					670								

-continued

Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	675	680	685
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	690	695	700
Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	705	710	715
Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	725	730	735
Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	740	745	750
Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	755	760	765
Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	770	775	780
Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	785	790	795
Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	805	810	815
Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	820	825	830
Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	835	840	845
Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	850	855	860
Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	865	870	875
Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	885	890	895
Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	900	905	910
Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	915	920	925
Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	930	935	940
Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	945	950	955
Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	965	970	975
Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	980	985	990
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	995	1000	1005
Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	1010	1015	1020
Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	1025	1030	1035
Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	1045	1050	1055
Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	1060	1065	1070

-continued

Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala
	1075						1080					1085			
Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu
	1090				1095						1100				
Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr
1105				1110						1115				1120	
Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn
			1125						1130					1135	

Phe

<210> SEQ ID NO 38
 <211> LENGTH: 1137
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 38

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

275																	280							285						
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn															
290						295				300																				
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala															
305				310						315				320																
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg															
				325				330						335																
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val															
		340						345				350																		
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val															
355						360						365																		
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp															
370						375				380																				
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu															
385				390						395				400																
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr															
				405				410						415																
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala															
		420						425				430																		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly															
435						440						445																		
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys															
450						455						460																		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp															
465				470						475				480																
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly															
				485				490						495																
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys															
		500						505				510																		
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn															
515						520						525																		
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val															
530						535				540																				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala															
545				550						555				560																
Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu															
		565						570						575																
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala															
		580						585				590																		
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg															
595						600						605																		
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val															
610						615				620																				
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val															
625				630						635				640																
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp															
				645				650						655																
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu															
		660						665				670																		
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr															
		675				680				685																				

Pro 690	Asp	Gln	Val	Val	Ala	Ile 695	Ala	Ser	Asn	Ile 700	Gly	Gly	Lys	Gln	Ala
Leu 705	Glu	Ser	Ile	Val	Ala 710	Gln	Leu	Ser	Arg	Arg 715	Asp	Pro	Ala	Leu	Ala 720
Ala	Leu	Thr	Asn	Asp 725	His	Leu	Val	Ala	Leu 730	Ala	Cys	Leu	Gly	Gly 735	Arg
Pro	Ala	Leu	Asp 740	Ala	Val	Lys	Lys	Gly 745	Leu	Pro	His	Ala	Pro 750	Glu	Phe
Ile	Arg 755	Arg	Val	Asn	Arg	Arg	Ile 760	Ala	Glu	Arg	Thr	Ser 765	His	Arg	Val
Ala 770	Asp	Tyr	Ala	His	Val	Val 775	Arg	Val	Leu	Glu	Phe 780	Phe	Gln	Cys	His
Ser 785	His	Pro	Ala	His	Ala 790	Phe	Asp	Glu	Ala	Met 795	Thr	Gln	Phe	Gly	Met 800
Ser	Arg	His	Gly 805	Leu	Val	Gln	Leu	Phe	Arg 810	Arg	Val	Gly	Val	Thr 815	Glu
Phe	Glu	Ala	Arg 820	Tyr	Gly	Thr	Leu	Pro 825	Pro	Ala	Ser	Gln	Arg 830	Trp	Asp
Arg	Ile 835	Leu	Gln	Ala	Ser	Gly	Met 840	Lys	Arg	Ala	Lys	Pro 845	Ser	Pro	Thr
Ser 850	Ala	Gln	Thr	Pro	Asp	Gln 855	Thr	Ser	Leu	His	Ala 860	Phe	Ala	Asp	Ser
Leu 865	Glu	Arg	Asp	Leu	Asp 870	Ala	Pro	Ser	Pro	Met 875	His	Glu	Gly	Asp	Gln 880
Thr	Arg	Ala	Ser 885	Ser	Arg	Lys	Arg	Ser 890	Arg	Ser	Asp	Arg	Ala	Val 895	Thr
Gly	Pro	Ser	Ala 900	Gln	Gln	Ala	Val	Glu 905	Val	Arg	Val	Pro 910	Glu	Gln	Arg
Asp	Ala 915	Leu	His	Leu	Pro	Leu	Ser 920	Trp	Arg	Val	Lys	Arg 925	Pro	Arg	Thr
Arg 930	Ile	Trp	Gly	Gly	Leu	Pro 935	Asp	Pro	Ile	Ser	Arg 940	Ser	Gln	Leu	Val
Lys 945	Ser	Glu	Leu	Glu	Glu 950	Lys	Lys	Ser	Glu	Leu 955	Arg	His	Lys	Leu	Lys 960
Tyr	Val	Pro	His 965	Glu	Tyr	Ile	Glu	Leu	Ile 970	Glu	Ile	Ala	Arg 975	Asn	Ser
Thr	Gln	Asp	Arg 980	Ile	Leu	Glu	Met	Lys 985	Val	Met	Glu	Phe 990	Phe	Met	Lys
Val	Tyr 995	Gly	Tyr	Arg	Gly	Lys	His 1000	Leu	Gly	Gly	Ser	Arg 1005	Lys	Pro	Asp
Gly 1010	Ala	Ile	Tyr	Thr	Val	Gly 1015	Ser	Pro	Ile	Asp	Tyr	Gly 1020	Val	Ile	Val
Asp 1025	Thr	Lys	Ala	Tyr	Ser 1030	Gly	Gly	Tyr	Asn	Leu 1035	Pro	Ile	Gly	Gln	Ala 1040
Asp	Glu	Met	Gln 1045	Arg	Tyr	Val	Glu	Glu	Asn 1050	Gln	Thr	Arg	Asn	Lys 1055	His
Ile	Asn	Pro	Asn 1060	Glu	Trp	Trp	Lys	Val 1065	Tyr	Pro	Ser	Ser 1070	Val	Thr	Glu
Phe	Lys 1075	Phe	Leu	Phe	Val	Ser	Gly 1080	His	Phe	Lys	Gly 1085	Asn	Tyr	Lys	Ala

-continued

Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu
1090 1095 1100

Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr
1105 1110 1115 1120

Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn
1125 1130 1135

Phe

<210> SEQ ID NO 39
 <211> LENGTH: 1137
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 39

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20 25 30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35 40 45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
50 55 60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
65 70 75 80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85 90 95

Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
100 105 110

Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
115 120 125

Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
130 135 140

Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
145 150 155 160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
165 170 175

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
180 185 190

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
195 200 205

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
210 215 220

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
225 230 235 240

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
245 250 255

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
260 265 270

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
275 280 285

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn

-continued

290	295	300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala 305 310 315 320		
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 325 330 335		
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 340 345 350		
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val 355 360 365		
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 370 375 380		
Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu 385 390 395 400		
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 405 410 415		
Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala 420 425 430		
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 435 440 445		
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys 450 455 460		
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 465 470 475 480		
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly 485 490 495		
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys 500 505 510		
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn 515 520 525		
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val 530 535 540		
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala 545 550 555 560		
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu 565 570 575		
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala 580 585 590		
Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 595 600 605		
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 610 615 620		
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val 625 630 635 640		
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 645 650 655		
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu 660 665 670		
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 675 680 685		
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala 690 695 700		

-continued

Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	
705					710					715					720	
Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	
			725						730					735		
Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	
			740					745					750			
Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	
	755						760					765				
Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	
	770					775					780					
Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	
785					790					795					800	
Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	
			805						810					815		
Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	
		820						825					830			
Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	
	835						840					845				
Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	
	850					855					860					
Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	
865					870					875					880	
Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	
			885						890					895		
Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	
			900						905					910		
Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	
	915						920					925				
Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	
	930					935					940					
Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	
945					950					955					960	
Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	
			965						970					975		
Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	
		980						985					990			
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	
	995					1000						1005				
Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	
	1010					1015					1020					
Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	
1025					1030					1035					1040	
Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	
			1045						1050					1055		
Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	
			1060						1065					1070		
Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	
	1075						1080					1085				
Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	
	1090					1095						1100				

-continued

Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr
1105 1110 1115 1120

Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn
1125 1130 1135

Phe

<210> SEQ ID NO 40

<211> LENGTH: 1137

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 40

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20 25 30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35 40 45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
50 55 60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
65 70 75 80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85 90 95

Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
100 105 110

Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
115 120 125

Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
130 135 140

Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
145 150 155 160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
165 170 175

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
180 185 190

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
195 200 205

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
210 215 220

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
225 230 235 240

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
245 250 255

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
260 265 270

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
275 280 285

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
290 295 300

Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala

-continued

305	310							315							320	
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
325				330				335								
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
340				345				350								
Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
355				360				365								
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
370				375				380								
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385				390				395				400				
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
405				410				415								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
420				425				430								
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
435				440				445								
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	
450				455				460								
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
465				470				475				480				
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
485				490				495								
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
500				505				510								
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
515				520				525								
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
530				535				540								
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
545				550				555				560				
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
565				570				575								
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
580				585				590								
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
595				600				605								
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
610				615				620								
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
625				630				635				640				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
645				650				655								
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
660				665				670								
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
675				680				685								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	
690				695				700								
Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	
705				710				715				720				

-continued

Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	
			725						730					735		
Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	
			740					745					750			
Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	
			755				760					765				
Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	
			770			775					780					
Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	
			785		790				795						800	
Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	
			805					810						815		
Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	
			820					825					830			
Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	
			835				840					845				
Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	
			850			855					860					
Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	
			865		870					875					880	
Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	
			885					890						895		
Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	
			900					905					910			
Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	
			915				920					925				
Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	
			930			935					940					
Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	
			945			950				955					960	
Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	
			965					970						975		
Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	
			980					985					990			
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	
			995			1000						1005				
Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	
			1010			1015					1020					
Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	
			1025		1030					1035					1040	
Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	
			1045					1050						1055		
Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	
			1060					1065						1070		
Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	
			1075				1080					1085				
Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	
			1090			1095					1100					
Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	
			1105		1110				1115						1120	

-continued

Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn
 1125 1130 1135

Phe

<210> SEQ ID NO 41
 <211> LENGTH: 1137
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 41

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
 1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
 20 25 30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
 35 40 45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
 50 55 60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
 65 70 75 80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
 85 90 95

Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
 100 105 110

Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
 115 120 125

Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
 130 135 140

Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
 145 150 155 160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
 165 170 175

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
 180 185 190

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
 195 200 205

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
 210 215 220

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
 225 230 235 240

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
 245 250 255

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
 260 265 270

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
 275 280 285

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
 290 295 300

Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala
 305 310 315 320

Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg

-continued

			325						330						335			
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val			
			340						345					350				
Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val			
			355				360						365					
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp			
	370						375					380						
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu			
385						390				395					400			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr			
			405						410					415				
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala			
			420						425					430				
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly			
		435					440					445						
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys			
	450						455					460						
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp			
465						470				475					480			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly			
			485						490					495				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys			
		500						505						510				
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn			
		515					520						525					
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val			
	530						535					540						
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala			
545					550					555					560			
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu			
			565						570					575				
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala			
		580						585					590					
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg			
		595					600						605					
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val			
		610					615					620						
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val			
625					630					635					640			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp			
			645						650					655				
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu			
		660						665					670					
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr			
		675					680						685					
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala			
	690						695					700						
Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala			
705					710					715					720			
Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg			
			725						730					735				

-continued

Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	740	745	750
Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	755	760	765
Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	770	775	780
Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	785	790	795
Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	805	810	815
Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	820	825	830
Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	835	840	845
Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	850	855	860
Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	865	870	875
Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	885	890	895
Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	900	905	910
Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	915	920	925
Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	930	935	940
Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	945	950	955
Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	965	970	975
Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	980	985	990
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	995	1000	1005
Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	1010	1015	1020
Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	1025	1030	1035
Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	1045	1050	1055
Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	1060	1065	1070
Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	1075	1080	1085
Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	1090	1095	1100
Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	1105	1110	1115
Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	1125	1130	1135

-continued

Phe

```

<210> SEQ ID NO 42
<211> LENGTH: 1171
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 42
Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1      5      10      15
Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20     25     30
Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35     40     45
Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
50     55     60
Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
65     70     75     80
Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85     90     95
Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
100    105    110
Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
115    120    125
Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
130    135    140
Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
145    150    155    160
Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
165    170    175
Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
180    185    190
Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
195    200    205
Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
210    215    220
Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
225    230    235    240
Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
245    250    255
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
260    265    270
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
275    280    285
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
290    295    300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala
305    310    315    320
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
325    330    335
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val

```

-continued

340							345					350				
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
355							360	365								
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
370							375	380								
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385							390	395					400			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
405							410	415								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
420							425	430								
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
435							440	445								
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
450							455	460								
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
465							470	475					480			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
485							490	495								
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
500							505	510								
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
515							520	525								
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
530							535	540								
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
545							550	555					560			
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
565							570	575								
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
580							585	590								
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
595							600	605								
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
610							615	620								
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
625							630	635					640			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
645							650	655								
Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
660							665	670								
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
675							680	685								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
690							695	700								
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
705							710	715					720			
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	
725							730	735								
Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	
740							745	750								

Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly
755															
Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro
770															
Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His
785															
Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln
805															
Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe
820															
Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val
835															
Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg
850															
Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser
865															
Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala
885															
Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly
900															
Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala
915															
Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu
930															
Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro
945															
Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln
965															
Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys
980															
Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg
995															
Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe
1010															
Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys
1025															
Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val
1045															
Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly
1060															
Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn
1075															
Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val
1090															
Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr
1105															
Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala
1125															
Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala
1140															

-continued

Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu
1155 1160 1165

Ile Asn Phe
1170

<210> SEQ ID NO 43
 <211> LENGTH: 1239
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 43

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

-continued

Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg		325	330	335
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val		340	345	350
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val		355	360	365
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp		370	375	380
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu		385	390	400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr		405	410	415
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala		420	425	430
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly		435	440	445
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys		450	455	460
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp		465	470	475
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly		485	490	495
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys		500	505	510
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn		515	520	525
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val		530	535	540
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala		545	550	555
Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu		565	570	575
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala		580	585	590
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg		595	600	605
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val		610	615	620
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val		625	630	635
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp		645	650	655
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu		660	665	670
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr		675	680	685
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala		690	695	700
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly		705	710	715
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys				

-continued

725								730					735				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp		
740								745					750				
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly		
755								760					765				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys		
770								775					780				
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn		
785								790					795				
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg		
805								810					815				
Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu		
820								825					830				
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu		
835								840					845				
Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu		
850								855					860				
Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu		
865								870					875				
Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala		
885								890					895				
Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg		
900								905					910				
Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro		
915								920					925				
Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg		
930								935					940				
Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu		
945								950					955				
His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro		
965								970					975				
Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg		
980								985					990				
Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val		
995								1000					1005				
Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg		
1010								1015					1020				
Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile		
1025								1030					1035				
Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu		
1045								1050					1055				
Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile		
1060								1065					1070				
Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val		
1075								1080					1085				
Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly		
1090								1095					1100				
Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile		
1105								1110					1115				
Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn		
1125								1130					1135				

-continued

Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn
 1140 1145 1150
 Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr
 1155 1160 1165
 Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe
 1170 1175 1180
 Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn
 1185 1190 1195 1200
 Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu
 1205 1210 1215
 Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe
 1220 1225 1230
 Asn Asn Gly Glu Ile Asn Phe
 1235

<210> SEQ ID NO 44
 <211> LENGTH: 1239
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 44

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

-continued

Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	
				245					250					255		
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	
			260					265					270			
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	
		275					280						285			
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	
	290						295				300					
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	
305					310					315					320	
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
				325					330						335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
			340					345						350		
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		355					360						365			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
	370						375				380					
Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385					390					395					400	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
				405					410						415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	
		420						425					430			
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
		435					440					445				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	
	450						455				460					
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
465					470					475					480	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
				485					490						495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		500						505					510			
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	
	515						520					525				
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
	530						535				540					
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
545					550					555					560	
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
				565					570						575	
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
		580						585					590			
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
		595					600						605			
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
	610					615					620					
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
625					630					635					640	

-continued

Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
				645					650					655		
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
			660					665					670			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
		675						680					685			
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	
	690					695					700					
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
705					710					715					720	
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
				725					730					735		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
			740					745					750			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
		755					760					765				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
	770					775					780					
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
785					790					795					800	
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	
				805					810					815		
Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	
			820					825					830			
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	
		835					840					845				
Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	
		850				855					860					
Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	
865					870					875					880	
Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	
				885					890					895		
Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	
		900						905					910			
Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	
		915					920					925				
Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	
		930				935						940				
Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	
945					950					955					960	
His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	
			965						970					975		
Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	
		980						985					990			
Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	
		995					1000					1005				
Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	
		1010				1015					1020					
Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	
1025					1030					1035					1040	
Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	

-continued

1045					1050					1055					
Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile
			1060					1065					1070		
Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val
		1075					1080					1085			
Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly
	1090						1095					1100			
Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile
	1105						1110					1115			1120
Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn
			1125						1130					1135	
Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn
			1140						1145				1150		
Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr
		1155					1160						1165		
Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe
	1170						1175					1180			
Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn
	1185						1190					1195			1200
Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu
			1205						1210				1215		
Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe
			1220					1225					1230		
Asn	Asn	Gly	Glu	Ile	Asn	Phe									
			1235												

<210> SEQ ID NO 45

<211> LENGTH: 1239

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 45

Met	Ala	Ser	Ser	Pro	Pro	Lys	Lys	Lys	Arg	Lys	Val	Ser	Trp	Lys	Asp
1				5					10					15	
Ala	Ser	Gly	Trp	Ser	Arg	Met	His	Ala	Asp	Pro	Ile	Arg	Pro	Arg	Arg
		20					25					30			
Pro	Ser	Pro	Ala	Arg	Glu	Leu	Leu	Pro	Gly	Pro	Gln	Pro	Asp	Arg	Val
		35				40					45				
Gln	Pro	Thr	Ala	Asp	Arg	Gly	Val	Ser	Ala	Pro	Ala	Gly	Ser	Pro	Leu
	50					55					60				
Asp	Gly	Leu	Pro	Ala	Arg	Arg	Thr	Val	Ser	Arg	Thr	Arg	Leu	Pro	Ser
	65				70				75					80	
Pro	Pro	Ala	Pro	Ser	Pro	Ala	Phe	Ser	Ala	Gly	Ser	Phe	Ser	Asp	Leu
			85					90						95	
Leu	Arg	Pro	Phe	Asp	Pro	Ser	Leu	Leu	Asp	Thr	Ser	Leu	Leu	Asp	Ser
		100					105					110			
Met	Pro	Ala	Val	Gly	Thr	Pro	His	Thr	Ala	Ala	Ala	Pro	Ala	Glu	Trp
	115					120						125			
Asp	Glu	Ala	Gln	Ser	Ala	Leu	Arg	Ala	Ala	Asp	Asp	Pro	Pro	Pro	Thr
	130					135					140				
Val	Arg	Val	Ala	Val	Thr	Ala	Ala	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala

-continued

145	150							155							160
Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln
				165					170					175	
Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile
			180					185					190		
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val
		195					200					205			
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro
	210					215					220				
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala
225					230					235					240
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp
				245					250					255	
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu
			260					265					270		
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala
		275					280						285		
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn
	290					295					300				
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala
305					310					315					320
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
			325						330					335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
		340						345					350		
Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
		355					360					365			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
	370					375					380				
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu
385					390					395					400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
			405						410					415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala
			420					425					430		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
		435					440					445			
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys
	450					455					460				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
465					470					475					480
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly
			485						490					495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
			500						505				510		
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn
		515					520					525			
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
	530					535					540				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
545					550					555					560

-continued

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

<400> SEQUENCE: 46

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

-continued

Asp	Gly	Leu	Pro	Ala	Arg	Arg	Thr	Val	Ser	Arg	Thr	Arg	Leu	Pro	Ser	65	70	75	80
Pro	Pro	Ala	Pro	Ser	Pro	Ala	Phe	Ser	Ala	Gly	Ser	Phe	Ser	Asp	Leu	85	90	95	
Leu	Arg	Pro	Phe	Asp	Pro	Ser	Leu	Leu	Asp	Thr	Ser	Leu	Leu	Asp	Ser	100	105	110	
Met	Pro	Ala	Val	Gly	Thr	Pro	His	Thr	Ala	Ala	Ala	Pro	Ala	Glu	Trp	115	120	125	
Asp	Glu	Ala	Gln	Ser	Ala	Leu	Arg	Ala	Ala	Asp	Asp	Pro	Pro	Pro	Thr	130	135	140	
Val	Arg	Val	Ala	Val	Thr	Ala	Ala	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala	145	150	155	160
Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln	165	170	175	
Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	180	185	190	
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	195	200	205	
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	210	215	220	
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	225	230	235	240
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	245	250	255	
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	260	265	270	
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	275	280	285	
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	290	295	300	
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	305	310	315	320
Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	325	330	335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	340	345	350	
Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	355	360	365	
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	370	375	380	
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	385	390	395	400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	405	410	415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	420	425	430	
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	435	440	445	
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	450	455	460	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp				

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

```
<210> SEQ ID NO 47
<211> LENGTH: 1239
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
```

-continued

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 47

```

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

```

-continued

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu	385	390	395	400
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	405	410	415	
Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala	420	425	430	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	435	440	445	
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys	450	455	460	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	465	470	475	480
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly	485	490	495	
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	500	505	510	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His	515	520	525	
Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	530	535	540	
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	545	550	555	560
Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	565	570	575	
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala	580	585	590	
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	595	600	605	
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	610	615	620	
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val	625	630	635	640
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	645	650	655	
Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu	660	665	670	
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	675	680	685	
Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala	690	695	700	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	705	710	715	720
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys	725	730	735	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	740	745	750	
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly	755	760	765	
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	770	775	780	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn				

-continued

785					790						795				800
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg
			805						810					815	
Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu
			820					825				830			
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu
		835					840					845			
Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu
	850					855				860					
Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu
865					870					875					880
Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala
			885						890					895	
Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg
		900						905					910		
Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro
		915					920					925			
Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg
	930					935					940				
Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu
945					950					955					960
His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro
			965						970					975	
Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg
		980						985					990		
Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val
	995						1000					1005			
Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg
	1010					1015					1020				
Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile
1025					1030					1035					1040
Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu
			1045						1050					1055	
Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile
		1060					1065						1070		
Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val
	1075						1080					1085			
Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly
	1090					1095					1100				
Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile
1105					1110					1115					1120
Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn
			1125						1130					1135	
Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn
		1140						1145					1150		
Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr
		1155					1160						1165		
Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe
	1170						1175					1180			
Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn
1185					1190					1195					1200

-continued

Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu
 1205 1210 1215

Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe
 1220 1225 1230

Asn Asn Gly Glu Ile Asn Phe
 1235

<210> SEQ ID NO 48
 <211> LENGTH: 1239
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 48

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
 1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
 20 25 30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
 35 40 45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
 50 55 60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
 65 70 75 80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
 85 90 95

Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser
 100 105 110

Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp
 115 120 125

Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr
 130 135 140

Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala
 145 150 155 160

Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln
 165 170 175

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
 180 185 190

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
 195 200 205

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
 210 215 220

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
 225 230 235 240

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
 245 250 255

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
 260 265 270

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
 275 280 285

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
 290 295 300

-continued

Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala
305					310					315					320
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
				325					330					335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
			340					345					350		
Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
	355						360					365			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
	370					375				380					
Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu
385					390				395						400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
				405				410						415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala
			420				425						430		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
	435					440				445					
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys
	450				455				460						
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
465					470				475						480
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly
				485				490						495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
		500					505					510			
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn
	515					520					525				
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
	530				535					540					
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
545				550					555						560
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
			565					570						575	
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
		580					585					590			
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
	595						600					605			
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
	610				615					620					
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
625				630					635						640
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
			645					650					655		
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu
		660					665					670			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
	675					680						685			
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala
	690				695						700				

-continued

Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
705					710					715					720	
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	
				725					730					735		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
			740					745					750			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	
	755						760					765				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
	770					775					780					
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
785					790					795					800	
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	
				805					810					815		
Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	
			820					825					830			
Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	Gly	Leu	
	835						840					845				
Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Ala	Glu	
	850					855					860					
Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	Val	Leu	
865					870					875					880	
Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	Glu	Ala	
				885					890					895		
Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	Phe	Arg	
		900						905					910			
Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	Pro	Pro	
	915						920					925				
Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	
	930					935					940					
Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	Ser	Leu	
945					950					955					960	
His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	
				965					970					975		
Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	
		980						985					990			
Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	
	995						1000					1005				
Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	
	1010					1015					1020					
Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	
1025					1030					1035					1040	
Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	
				1045					1050					1055		
Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	
		1060						1065					1070			
Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	
		1075					1080					1085				
Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	
	1090					1095					1100					
Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	

-continued

1105	1110	1115	1120
Asp Tyr Gly Val	Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn		
	1125	1130	1135
Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn			
	1140	1145	1150
Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr			
	1155	1160	1165
Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe			
	1170	1175	1180
Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn			
	1185	1190	1195
Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu			
	1205	1210	1215
Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe			
	1220	1225	1230
Asn Asn Gly Glu Ile Asn Phe			
	1235		

<210> SEQ ID NO 49
 <211> LENGTH: 1273
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 49

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

-continued

210	215	220
Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala		
225	230	235 240
Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp		
	245	250 255
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu		
	260	265 270
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala		
	275	280 285
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn		
	290	295 300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala		
305	310	315 320
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
	325	330 335
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		
	340	345 350
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val		
	355	360 365
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp		
	370	375 380
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu		
385	390	395 400
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr		
	405	410 415
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala		
	420	425 430
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
	435	440 445
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys		
	450	455 460
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
465	470	475 480
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly		
	485	490 495
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
	500	505 510
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His		
	515	520 525
Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val		
	530	535 540
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
545	550	555 560
Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
	565	570 575
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala		
	580	585 590
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
	595	600 605
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		
	610	615 620

-continued

Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	625	630	635	640
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	645	650	655	
Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	660	665	670	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	675	680	685	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	690	695	700	
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	705	710	715	720
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	725	730	735	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	740	745	750	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	755	760	765	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	770	775	780	
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	785	790	795	800
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	805	810	815	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	820	825	830	
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	835	840	845	
Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	850	855	860	
Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys	865	870	875	880
Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	885	890	895	
Ala	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg	900	905	910	
Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp	915	920	925	
Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu	930	935	940	
Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu	945	950	955	960
Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	965	970	975	
Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr	980	985	990	
Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	995	1000	1005	
Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	1010	1015	1020	

-continued

```

Ser Arg Ser Asp Arg Ala Val Thr Gly Pro Ser Ala Gln Gln Ala Val
1025                      1030                      1035                      1040

Glu Val Arg Val Pro Glu Gln Arg Asp Ala Leu His Leu Pro Leu Ser
                      1045                      1050                      1055

Trp Arg Val Lys Arg Pro Arg Thr Arg Ile Trp Gly Gly Leu Pro Asp
                      1060                      1065                      1070

Pro Ile Ser Arg Ser Gln Leu Val Lys Ser Glu Leu Glu Glu Lys Lys
                      1075                      1080                      1085

Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu
                      1090                      1095                      1100

Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu Met
1105                      1110                      1115                      1120

Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys His
                      1125                      1130                      1135

Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser
                      1140                      1145                      1150

Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly
                      1155                      1160                      1165

Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu
                      1170                      1175                      1180

Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys
1185                      1190                      1195                      1200

Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly
                      1205                      1210                      1215

His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile
                      1220                      1225                      1230

Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly
                      1235                      1240                      1245

Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg
                      1250                      1255                      1260

Lys Phe Asn Asn Gly Glu Ile Asn Phe
1265                      1270

```

```

<210> SEQ ID NO 50
<211> LENGTH: 1273
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

```

```

<400> SEQUENCE: 50

```

```

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1      5      10      15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20     25     30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35     40     45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu
50     55     60

Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser
65     70     75     80

Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu
85     90     95

```

-continued

Leu	Arg	Pro	Phe	Asp	Pro	Ser	Leu	Leu	Asp	Thr	Ser	Leu	Leu	Asp	Ser	
			100					105					110			
Met	Pro	Ala	Val	Gly	Thr	Pro	His	Thr	Ala	Ala	Ala	Pro	Ala	Glu	Trp	
		115					120					125				
Asp	Glu	Ala	Gln	Ser	Ala	Leu	Arg	Ala	Ala	Asp	Asp	Pro	Pro	Pro	Thr	
	130					135					140					
Val	Arg	Val	Ala	Val	Thr	Ala	Ala	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala	
145					150					155					160	
Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln	
			165						170					175		
Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	
	180							185					190			
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	
	195						200					205				
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	
	210					215					220					
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	
225					230					235					240	
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	
			245					250						255		
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	
		260						265					270			
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	
	275						280					285				
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	
	290					295					300					
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	
305					310					315					320	
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
			325						330					335		
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
		340						345					350			
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
	355						360					365				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
	370					375					380					
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385					390					395					400	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			405						410					415		
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	
		420					425					430				
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
	435					440						445				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	
	450					455					460					
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
465					470					475					480	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
		485							490					495		
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	

-continued

500							505					510				
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
515							520					525				
Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
530							535					540				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
545							550					555				
Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
565							570					575				
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
580							585					590				
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
595							600					605				
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
610							615					620				
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
625							630					635				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
645							650					655				
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
660							665					670				
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
675							680					685				
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	
690							695					700				
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
705							710					715				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	
725							730					735				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
740							745					750				
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	
755							760					765				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
770							775					780				
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
785							790					795				
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
805							810					815				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
820							825					830				
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	
835							840					845				
Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	
850							855					860				
Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	Lys	Lys	
865							870					875				
Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	
885							890					895				
Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	Gln	Val	Val	Arg	
900							905					910				

-continued

Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	Tyr	Ala	Phe	Asp	915	920	925
Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	Asn	Gly	Leu	Val	Gln	Leu	930	935	940
Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Leu	Glu	Ala	Arg	Gly	Gly	Thr	Leu	945	950	955
Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	965	970	975
Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Ala	980	985	990
Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	995	1000	1005
Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	1010	1015	1020
Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	1025	1030	1035
Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	1045	1050	1055
Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	1060	1065	1070
Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	1075	1080	1085
Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	1090	1095	1100
Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	1105	1110	1115
Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	1125	1130	1135
Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	1140	1145	1150
Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	1155	1160	1165
Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	1170	1175	1180
Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	1185	1190	1195
Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	1205	1210	1215
His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	1220	1225	1230
Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	1235	1240	1245
Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	1250	1255	1260
Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe								1265	1270	

<210> SEQ ID NO 51

<211> LENGTH: 1273

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 51

```

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

```

-continued

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu	385	390	395	400
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	405	410	415	
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	420	425	430	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	435	440	445	
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	450	455	460	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	465	470	475	480
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly	485	490	495	
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	500	505	510	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His	515	520	525	
Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	530	535	540	
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	545	550	555	560
Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	565	570	575	
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala	580	585	590	
Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	595	600	605	
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	610	615	620	
Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val	625	630	635	640
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	645	650	655	
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu	660	665	670	
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	675	680	685	
Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala	690	695	700	
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	705	710	715	720
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys	725	730	735	
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	740	745	750	
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly	755	760	765	
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	770	775	780	
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His				

-continued

785					790						795				800
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
				805					810					815	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
			820					825					830		
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu
		835					840					845			
Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val
	850					855					860				
Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	Lys	Lys
865					870					875				880	
Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile
			885						890					895	
Ala	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	Val	Arg
			900					905					910		
Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	Phe	Asp
		915					920					925			
Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	Gln	Leu
	930					935					940				
Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	Thr	Leu
945					950					955				960	
Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met
			965					970						975	
Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	Gln	Thr
		980						985					990		
Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro
	995						1000					1005			
Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg
	1010					1015					1020				
Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val
1025					1030					1035				1040	
Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	Leu	Ser
			1045						1050					1055	
Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp
		1060						1065					1070		
Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys
	1075					1080						1085			
Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu
	1090					1095					1100				
Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met
1105				1110						1115				1120	
Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His
			1125						1130					1135	
Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser
		1140						1145					1150		
Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly
	1155					1160						1165			
Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu
	1170					1175					1180				
Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys
1185					1190					1195				1200	

```
<210> SEQ ID NO 52
<211> LENGTH: 1307
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 52
```

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

-continued

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala	275	280	285
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn	290	295	300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala	305	310	315
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	325	330	335
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	340	345	350
Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val	355	360	365
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	370	375	380
Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu	385	390	395
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr	405	410	415
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala	420	425	430
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly	435	440	445
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys	450	455	460
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp	465	470	475
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly	485	490	495
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	500	505	510
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn	515	520	525
Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	530	535	540
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	545	550	555
Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	565	570	575
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala	580	585	590
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	595	600	605
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	610	615	620
Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val	625	630	635
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp	645	650	655
Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu	660	665	670

-continued

Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	675	680	685
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	690	695	700
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	705	710	715
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	725	730	735
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	740	745	750
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	755	760	765
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	770	775	780
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	785	790	795
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	805	810	815
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	820	825	830
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	835	840	845
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	850	855	860
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	865	870	875
Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	885	890	895
Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	Ala	Val	900	905	910
Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	Asn	Arg	915	920	925
Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	His	Val	930	935	940
Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	His	Ala	945	950	955
Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	Leu	Val	965	970	975
Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	Tyr	Gly	980	985	990
Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	995	1000	1005
Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	1010	1015	1020
Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	1025	1030	1035
Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	1045	1050	1055
Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	1060	1065	1070
Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro			

-continued

1075	1080	1085
Leu Ser Trp Arg Val Lys Arg Pro Arg Thr Arg Ile Trp Gly Gly Leu 1090 1095 1100		
Pro Asp Pro Ile Ser Arg Ser Gln Leu Val Lys Ser Glu Leu Glu Glu 1105 1110 1115 1120		
Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr 1125 1130 1135		
Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu 1140 1145 1150		
Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly 1155 1160 1165		
Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val 1170 1175 1180		
Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser 1185 1190 1195 1200		
Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr 1205 1210 1215		
Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp 1220 1225 1230		
Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val 1235 1240 1245		
Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn 1250 1255 1260		
His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu 1265 1270 1275 1280		
Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val 1285 1290 1295		
Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe 1300 1305		

<210> SEQ ID NO 53

<211> LENGTH: 1341

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 53

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

-continued

115	120	125
Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr 130 135 140		
Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala 145 150 155 160		
Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln 165 170 175		
Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile 180 185 190		
Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val 195 200 205		
Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro 210 215 220		
Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala 225 230 235 240		
Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp 245 250 255		
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu 260 265 270		
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala 275 280 285		
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn 290 295 300		
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala 305 310 315 320		
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg 325 330 335		
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val 340 345 350		
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val 355 360 365		
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 370 375 380		
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu 385 390 395 400		
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr 405 410 415		
Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala 420 425 430		
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly 435 440 445		
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys 450 455 460		
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp 465 470 475 480		
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly 485 490 495		
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys 500 505 510		
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn 515 520 525		

Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
530			535			540			545			550			
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
545			550			555			560			565			
Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
565			570			575			580			585			
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
580			585			590			595			600			
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
595			600			605			610			615			
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
610			615			620			625			630			
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
625			630			635			640			645			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
645			650			655			660			665			
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu
660			665			670			675			680			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
675			680			685			690			695			
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala
690			695			700			705			710			
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
705			710			715			720			725			
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys
725			730			735			740			745			
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
740			745			750			755			760			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly
755			760			765			770			775			
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
770			775			780			785			790			
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His
785			790			795			800			805			
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
805			810			815			820			825			
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
820			825			830			835			840			
Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
835			840			845			850			855			
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
850			855			860			865			870			
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
865			870			875			880			885			
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
885			890			895			900			905			
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile
900			905			910			915			920			
Val	Ala	Gln	Leu	Ser	Arg	Arg	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn
915			920			925			930						

-continued

Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Leu	Asp	930	935	940
Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Phe	Ile	Arg	Arg	Val	945	950	955
Asn	Arg	Arg	Ile	Ala	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	965	970	975
His	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	980	985	990
His	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	His	Gly	995	1000	1005
Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Phe	Glu	Ala	Arg	1010	1015	1020
Tyr	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	1025	1030	1035
Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	1045	1050	1055
Pro	Asp	Gln	Thr	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	1060	1065	1070
Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	1075	1080	1085
Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	1090	1095	1100
Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	1105	1110	1115
Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	1125	1130	1135
Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	1140	1145	1150
Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	1155	1160	1165
Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	1170	1175	1180
Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	1185	1190	1195
Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	1205	1210	1215
Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	1220	1225	1230
Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	1235	1240	1245
Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	1250	1255	1260
Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	1265	1270	1275
Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	1285	1290	1295
Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	1300	1305	1310
Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	Leu	Thr	Leu	Glu	1315	1320	1325
Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	Phe						

-continued

1330	1335	1340
<210> SEQ ID NO 54		
<211> LENGTH: 1341		
<212> TYPE: PRT		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: synthetic peptide		
<400> SEQUENCE: 54		
Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp		
1	5	10 15
Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg		
	20	25 30
Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val		
	35	40 45
Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu		
	50	55 60
Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser		
65	70	75 80
Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu		
	85	90 95
Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser		
	100	105 110
Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp		
	115	120 125
Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr		
	130	135 140
Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala		
145	150	155 160
Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln		
	165	170 175
Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile		
	180	185 190
Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val		
	195	200 205
Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro		
	210	215 220
Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala		
225	230	235 240
Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp		
	245	250 255
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu		
	260	265 270
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala		
	275	280 285
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn		
	290	295 300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala		
305	310	315 320
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
	325	330 335
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		

-continued

340							345					350				
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
355							360	365								
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
370							375	380								
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
385							390	395					400			
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
405							410	415								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
420							425	430								
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
435							440	445								
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
450							455	460								
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
465							470	475					480			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	
485							490	495								
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
500							505	510								
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
515							520	525								
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
530							535	540								
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
545							550	555					560			
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
565							570	575								
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
580							585	590								
Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
595							600	605								
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
610							615	620								
Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
625							630	635					640			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
645							650	655								
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
660							665	670								
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
675							680	685								
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	
690							695	700								
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
705							710	715					720			
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	
725							730	735								
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
740							745	750								

-continued

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly	755	760	765
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys	770	775	780
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn	785	790	800
Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val	805	810	815
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala	820	825	830
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu	835	840	845
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala	850	855	860
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg	865	870	875
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val	885	890	895
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Ser Ile	900	905	910
Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn	915	920	925
Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp	930	935	940
Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val	945	950	955
Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Asp Tyr Ala	965	970	975
Gln Val Val Arg Val Leu Glu Phe Phe Gln Cys His Ser His Pro Ala	980	985	990
Tyr Ala Phe Asp Glu Ala Met Thr Gln Phe Gly Met Ser Arg Asn Gly	995	1000	1005
Leu Val Gln Leu Phe Arg Arg Val Gly Val Thr Glu Leu Glu Ala Arg	1010	1015	1020
Gly Gly Thr Leu Pro Pro Ala Ser Gln Arg Trp Asp Arg Ile Leu Gln	1025	1030	1035
Ala Ser Gly Met Lys Arg Ala Lys Pro Ser Pro Thr Ser Ala Gln Thr	1045	1050	1055
Pro Asp Gln Ala Ser Leu His Ala Phe Ala Asp Ser Leu Glu Arg Asp	1060	1065	1070
Leu Asp Ala Pro Ser Pro Met His Glu Gly Asp Gln Thr Arg Ala Ser	1075	1080	1085
Ser Arg Lys Arg Ser Arg Ser Asp Arg Ala Val Thr Gly Pro Ser Ala	1090	1095	1100
Gln Gln Ala Val Glu Val Arg Val Pro Glu Gln Arg Asp Ala Leu His	1105	1110	1115
Leu Pro Leu Ser Trp Arg Val Lys Arg Pro Arg Thr Arg Ile Trp Gly	1125	1130	1135
Gly Leu Pro Asp Pro Ile Ser Arg Ser Gln Leu Val Lys Ser Glu Leu	1140	1145	1150

-continued

Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His
 1155 1160 1165
 Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg
 1170 1175 1180
 Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr
 1185 1190 1195 1200
 Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr
 1205 1210 1215
 Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala
 1220 1225 1230
 Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln
 1235 1240 1245
 Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn
 1250 1255 1260
 Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu
 1265 1270 1275 1280
 Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg
 1285 1290 1295
 Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu
 1300 1305 1310
 Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu
 1315 1320 1325
 Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
 1330 1335 1340

<210> SEQ ID NO 55
 <211> LENGTH: 1341
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 55

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

-continued

Pro	Arg	Arg	Arg	Ala	Ala	Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln	
				165					170					175		
Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	
			180					185					190			
Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	
		195					200					205				
Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	
	210					215					220					
Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	
	225				230					235					240	
Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	
			245						250					255		
Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	
			260					265					270			
Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	
		275					280					285				
Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	
	290					295					300					
Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	
	305				310					315					320	
Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
			325					330						335		
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
			340					345					350			
Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		355					360					365				
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
	370					375					380					
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
	385				390					395					400	
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			405					410					415			
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	
			420					425					430			
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
		435					440					445				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	
	450					455					460					
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
	465				470					475					480	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
			485						490					495		
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		500						505					510			
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
		515					520					525				
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
	530					535					540					
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
	545				550					555					560	
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	

-continued

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

-continued

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

<210> SEQ ID NO 56

<211> LENGTH: 51

<212> TYPE: DNA

<213> ORGANISM: Arabidopsis sp.

-continued

<400> SEQUENCE: 56

tatcaagatt ctcttcactt ctctctgtca caccgatggt tacttctggg a 51

<210> SEQ ID NO 57

<211> LENGTH: 50

<212> TYPE: DNA

<213> ORGANISM: Arabidopsis sp.

<400> SEQUENCE: 57

tccggatgct cctcttgaca aggtctgtat tgtcagttgt ggtttgtcta 50

<210> SEQ ID NO 58

<211> LENGTH: 48

<212> TYPE: DNA

<213> ORGANISM: Arabidopsis sp.

<400> SEQUENCE: 58

ccggatgctc ctcttgacaa ggtctgtatt gtcagttgtg gtttgtct 48

<210> SEQ ID NO 59

<211> LENGTH: 40

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 59

ccggatgctc ctcttgacaa ttgtcagttg tggttgtct 40

<210> SEQ ID NO 60

<211> LENGTH: 43

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 60

ccggatgctc ctcttgacaa gtattgtcag ttgtggttg tct 43

<210> SEQ ID NO 61

<211> LENGTH: 33

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 61

ccggatgctc ctcttgacaa ttgtggttg tct 33

<210> SEQ ID NO 62

<211> LENGTH: 43

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 62

ccggatgctc ctcttgacaa ggattgtcag ttgtggttg tct 43

<210> SEQ ID NO 63

<211> LENGTH: 41

<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 63

ccggatgctc ctcttgacaa attgtcagtt gtggtttgtc t 41

<210> SEQ ID NO 64

<211> LENGTH: 44

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: modified Arabidopsis

<400> SEQUENCE: 64

ccggatgctc ctcttgacaa ggtattgtca gttgtggttt gtct 44

<210> SEQ ID NO 65

<211> LENGTH: 34

<212> TYPE: PRT

<213> ORGANISM: Xanthomonas gardneri

<400> SEQUENCE: 65

Leu Asp Thr Gly Gln Leu Phe Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 66

<211> LENGTH: 34

<212> TYPE: PRT

<213> ORGANISM: Xanthomonas campestris

<400> SEQUENCE: 66

Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 67

<211> LENGTH: 34

<212> TYPE: PRT

<213> ORGANISM: Xanthomonas oryzae

<400> SEQUENCE: 67

Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15Ala Val Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 68

<211> LENGTH: 34

<212> TYPE: PRT

<213> ORGANISM: Xanthomonas citri

<400> SEQUENCE: 68

-continued

Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15

Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 69
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 69

Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15

Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 70
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 70

Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr
1 5 10 15

Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
20 25 30

Leu Asn

<210> SEQ ID NO 71
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: *Xanthomonas oryzae*

<400> SEQUENCE: 71

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala
20 25 30

His Gly

<210> SEQ ID NO 72
<211> LENGTH: 1307
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 72

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp
1 5 10 15

Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg
20 25 30

Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val
35 40 45

Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu

50					55					60						
Asp 65	Gly	Leu	Pro	Ala	Arg 70	Arg	Thr	Val	Ser	Arg 75	Thr	Arg	Leu	Pro	Ser 80	
Pro	Pro	Ala	Pro	Ser 85	Pro	Ala	Phe	Ser	Ala 90	Gly	Ser	Phe	Ser	Asp 95	Leu	
Leu	Arg	Pro	Phe 100	Asp	Pro	Ser	Leu	Leu 105	Asp	Thr	Ser	Leu	Leu 110	Asp	Ser	
Met	Pro	Ala 115	Val	Gly	Thr	Pro	His 120	Thr	Ala	Ala	Ala	Pro 125	Ala	Glu	Trp	
Asp	Glu 130	Ala	Gln	Ser	Ala 135	Leu	Arg	Ala	Ala	Asp 140	Asp	Pro	Pro	Pro	Thr	
Val 145	Arg	Val	Ala	Val	Thr 150	Ala	Ala	Arg	Pro	Pro 155	Arg	Ala	Lys	Pro	Ala	
Pro	Arg	Arg	Arg	Ala 165	Ala	Gln	Pro	Ser	Asp 170	Ala	Ser	Pro	Ala	Ala 175	Gln	
Val	Asp	Leu	Arg 180	Thr	Leu	Gly	Tyr	Ser 185	Gln	Gln	Gln	Gln	Glu 190	Lys	Ile	
Lys	Pro	Lys 195	Val	Arg	Ser	Thr	Val 200	Ala	Gln	His	His	Glu 205	Ala	Leu	Val	
Gly	His 210	Gly	Phe	Thr	His	Ala 215	His	Ile	Val	Ala	Leu	Ser 220	Gln	His	Pro	
Ala 225	Ala	Leu	Gly	Thr	Val 230	Ala	Val	Thr	Tyr	Gln 235	His	Ile	Ile	Thr	Ala 240	
Leu	Pro	Glu	Ala	Thr 245	His	Glu	Asp	Ile	Val 250	Gly	Val	Gly	Lys	Gln 255	Trp	
Ser	Gly	Ala	Arg 260	Ala	Leu	Glu	Ala	Leu 265	Leu	Thr	Asp	Ala	Gly 270	Glu	Leu	
Arg	Gly	Pro 275	Pro	Leu	Gln	Leu	Asp 280	Thr	Gly	Gln	Leu	Val 285	Lys	Ile	Ala	
Lys	Arg 290	Gly	Gly	Val	Thr	Ala 295	Met	Glu	Ala	Val	His 300	Ala	Ser	Arg	Asn	
Ala 305	Leu	Thr	Gly	Ala	Pro 310	Leu	Asn	Leu	Thr	Pro 315	Ala	Gln	Val	Val	Ala 320	
Ile	Ala	Ser	Asn	Asn 325	Gly	Gly	Lys	Gln	Ala 330	Leu	Glu	Thr	Val	Gln 335	Arg	
Leu	Leu	Pro	Val 340	Leu	Cys	Gln	Asp	His 345	Gly	Leu	Thr	Pro 350	Asp	Gln	Val	
Val	Ala	Ile 355	Ala	Ser	His	Asp	Gly 360	Gly	Lys	Gln	Ala	Leu 365	Glu	Thr	Val	
Gln	Arg 370	Leu	Leu	Pro	Val	Leu 375	Cys	Gln	Asp	His 380	Gly	Leu	Thr	Pro	Asp	
Gln 385	Val	Val	Ala	Ile	Ala 390	Ser	Asn	Ile	Gly	Gly 395	Lys	Gln	Ala	Leu	Glu 400	
Thr	Val	Gln	Arg	Leu	Leu 405	Pro	Val	Leu	Cys	Gln 410	Asp	His	Gly	Leu	Thr 415	
Pro	Asp	Gln	Val 420	Val	Ala	Ile	Ala	Ser	Asn 425	Gly	Gly	Gly	Lys 430	Gln	Ala	
Leu	Glu	Thr	Val 435	Gln	Arg	Leu	Leu 440	Pro	Val	Leu	Cys	Gln 445	Asp	His	Gly	
Leu	Thr 450	Pro	Asp	Gln	Val	Val 455	Ala	Ile	Ala	Ser	Asn 460	Ile	Gly	Gly	Lys	

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

-continued

Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	
865					870					875					880	
Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	
				885					890					895		
Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met	Asp	Ala	Val	
			900					905					910			
Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	Arg	Val	Asn	Arg	
		915					920					925				
Arg	Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr	Ala	Gln	Val	
	930					935					940					
Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro	Ala	Tyr	Ala	
945					950					955					960	
Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Arg	Asn	Gly	Leu	Val	
				965					970					975		
Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	Leu	Glu	Ala	Arg	Gly	Gly	
		980						985					990			
Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	Arg	Ile	Leu	Gln	Ala	Ser	
	995					1000						1005				
Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	Ser	Ala	Gln	Thr	Pro	Asp	
	1010					1015					1020					
Gln	Ala	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	Leu	Glu	Arg	Asp	Leu	Asp	
1025					1030					1035				1040		
Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	Thr	Arg	Ala	Ser	Ser	Arg	
				1045					1050					1055		
Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	Gly	Pro	Ser	Ala	Gln	Gln	
		1060						1065					1070			
Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	Asp	Ala	Leu	His	Leu	Pro	
	1075						1080					1085				
Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	Arg	Ile	Trp	Gly	Gly	Leu	
	1090				1095						1100					
Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	Lys	Ser	Glu	Leu	Glu	Glu	
1105					1110					1115				1120		
Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	Tyr	Val	Pro	His	Glu	Tyr	
			1125						1130					1135		
Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	Thr	Gln	Asp	Arg	Ile	Leu	
		1140					1145						1150			
Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	Val	Tyr	Gly	Tyr	Arg	Gly	
	1155					1160						1165				
Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	Gly	Ala	Ile	Tyr	Thr	Val	
	1170					1175					1180					
Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	Asp	Thr	Lys	Ala	Tyr	Ser	
1185					1190					1195				1200		
Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	Asp	Glu	Met	Gln	Arg	Tyr	
			1205					1210					1215			
Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	Ile	Asn	Pro	Asn	Glu	Trp	
		1220					1225						1230			
Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	Phe	Lys	Phe	Leu	Phe	Val	
		1235					1240					1245				
Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	Gln	Leu	Thr	Arg	Leu	Asn	
	1250					1255					1260					
His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	Ser	Val	Glu	Glu	Leu	Leu	

-continued

1265	1270	1275	1280
Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val			
	1285	1290	1295
Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe			
	1300	1305	

<210> SEQ ID NO 73
 <211> LENGTH: 1409
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 73

Met Ala Ser Ser Pro Pro Lys Lys Lys Arg Lys Val Ser Trp Lys Asp	
1	15
Ala Ser Gly Trp Ser Arg Met His Ala Asp Pro Ile Arg Pro Arg Arg	
20	30
Pro Ser Pro Ala Arg Glu Leu Leu Pro Gly Pro Gln Pro Asp Arg Val	
35	45
Gln Pro Thr Ala Asp Arg Gly Val Ser Ala Pro Ala Gly Ser Pro Leu	
50	60
Asp Gly Leu Pro Ala Arg Arg Thr Val Ser Arg Thr Arg Leu Pro Ser	
65	80
Pro Pro Ala Pro Ser Pro Ala Phe Ser Ala Gly Ser Phe Ser Asp Leu	
85	95
Leu Arg Pro Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu Leu Asp Ser	
100	110
Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro Ala Glu Trp	
115	125
Asp Glu Ala Gln Ser Ala Leu Arg Ala Ala Asp Asp Pro Pro Pro Thr	
130	140
Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala Lys Pro Ala	
145	160
Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala Gln	
165	175
Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile	
180	190
Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val	
195	205
Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro	
210	220
Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala	
225	240
Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp	
245	255
Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu	
260	270
Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala	
275	285
Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn	
290	300
Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Ala Gln Val Val Ala	

-continued

305					310					315					320
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
				325					330					335	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
			340					345					350		
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
		355					360					365			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
	370					375					380				
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu
385					390				395						400
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
			405					410						415	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala
			420					425					430		
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
	435					440					445				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Asn	Gly	Gly	Lys
	450					455					460				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
465					470				475						480
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly
			485					490						495	
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
		500						505					510		
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn
	515					520						525			
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
	530					535					540				
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
545				550					555						560
Ser	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
			565					570						575	
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
		580						585					590		
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
		595					600						605		
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
	610					615					620				
Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
625					630				635						640
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
			645					650						655	
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu
			660					665					670		
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
		675					680					685			
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala
	690					695					700				
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
705					710					715					720

-continued

Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	
				725					730					735		
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
			740					745					750			
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	
		755					760					765				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
	770					775					780					
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	
	785				790					795					800	
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
				805					810						815	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
		820						825					830			
Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
		835					840					845				
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	
	850					855					860					
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	
	865				870					875					880	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	
				885					890						895	
Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	
		900					905						910			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	
	915						920					925				
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	
	930						935				940					
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
	945				950					955					960	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	
			965						970					975		
Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp	Pro	Ala	Leu	Ala	
		980						985					990			
Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	
		995					1000					1005				
Pro	Ala	Met	Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	
	1010						1015				1020					
Ile	Arg	Arg	Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	
	1025				1030					1035					1040	
Ala	Asp	Tyr	Ala	Gln	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	
			1045						1050					1055		
Ser	His	Pro	Ala	Tyr	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	
			1060					1065					1070			
Ser	Arg	Asn	Gly	Leu	Val	Gln	Leu	Phe	Arg	Arg	Val	Gly	Val	Thr	Glu	
		1075					1080					1085				
Leu	Glu	Ala	Arg	Gly	Gly	Thr	Leu	Pro	Pro	Ala	Ser	Gln	Arg	Trp	Asp	
	1090					1095					1100					
Arg	Ile	Leu	Gln	Ala	Ser	Gly	Met	Lys	Arg	Ala	Lys	Pro	Ser	Pro	Thr	
	1105					1110				1115					1120	

-continued

Ser	Ala	Gln	Thr	Pro	Asp	Gln	Ala	Ser	Leu	His	Ala	Phe	Ala	Asp	Ser	
				1125					1130					1135		
Leu	Glu	Arg	Asp	Leu	Asp	Ala	Pro	Ser	Pro	Met	His	Glu	Gly	Asp	Gln	
			1140					1145					1150			
Thr	Arg	Ala	Ser	Ser	Arg	Lys	Arg	Ser	Arg	Ser	Asp	Arg	Ala	Val	Thr	
		1155					1160					1165				
Gly	Pro	Ser	Ala	Gln	Gln	Ala	Val	Glu	Val	Arg	Val	Pro	Glu	Gln	Arg	
	1170					1175					1180					
Asp	Ala	Leu	His	Leu	Pro	Leu	Ser	Trp	Arg	Val	Lys	Arg	Pro	Arg	Thr	
	1185				1190					1195					1200	
Arg	Ile	Trp	Gly	Gly	Leu	Pro	Asp	Pro	Ile	Ser	Arg	Ser	Gln	Leu	Val	
			1205						1210					1215		
Lys	Ser	Glu	Leu	Glu	Glu	Lys	Lys	Ser	Glu	Leu	Arg	His	Lys	Leu	Lys	
		1220						1225					1230			
Tyr	Val	Pro	His	Glu	Tyr	Ile	Glu	Leu	Ile	Glu	Ile	Ala	Arg	Asn	Ser	
		1235					1240					1245				
Thr	Gln	Asp	Arg	Ile	Leu	Glu	Met	Lys	Val	Met	Glu	Phe	Phe	Met	Lys	
	1250					1255					1260					
Val	Tyr	Gly	Tyr	Arg	Gly	Lys	His	Leu	Gly	Gly	Ser	Arg	Lys	Pro	Asp	
	1265				1270					1275					1280	
Gly	Ala	Ile	Tyr	Thr	Val	Gly	Ser	Pro	Ile	Asp	Tyr	Gly	Val	Ile	Val	
			1285						1290					1295		
Asp	Thr	Lys	Ala	Tyr	Ser	Gly	Gly	Tyr	Asn	Leu	Pro	Ile	Gly	Gln	Ala	
		1300						1305					1310			
Asp	Glu	Met	Gln	Arg	Tyr	Val	Glu	Glu	Asn	Gln	Thr	Arg	Asn	Lys	His	
		1315					1320					1325				
Ile	Asn	Pro	Asn	Glu	Trp	Trp	Lys	Val	Tyr	Pro	Ser	Ser	Val	Thr	Glu	
	1330					1335					1340					
Phe	Lys	Phe	Leu	Phe	Val	Ser	Gly	His	Phe	Lys	Gly	Asn	Tyr	Lys	Ala	
	1345				1350					1355					1360	
Gln	Leu	Thr	Arg	Leu	Asn	His	Ile	Thr	Asn	Cys	Asn	Gly	Ala	Val	Leu	
			1365						1370					1375		
Ser	Val	Glu	Glu	Leu	Leu	Ile	Gly	Gly	Glu	Met	Ile	Lys	Ala	Gly	Thr	
		1380					1385					1390				
Leu	Thr	Leu	Glu	Glu	Val	Arg	Arg	Lys	Phe	Asn	Asn	Gly	Glu	Ile	Asn	
		1395					1400					1405				

Phe

<210> SEQ ID NO 74
 <211> LENGTH: 8
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 74

Lys Ile Ala Lys Arg Gly Gly Val
 1 5

<210> SEQ ID NO 75
 <211> LENGTH: 8
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic peptide

-continued

<400> SEQUENCE: 75

Lys Ile Ala Asn Gly Gly Gly Val
1 5

<210> SEQ ID NO 76

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 76

Lys Ile Ala Asn Ile Gly Gly Val
1 5

<210> SEQ ID NO 77

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 77

Lys Ile Ala His Asp Gly Gly Val
1 5

<210> SEQ ID NO 78

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 78

Lys Ile Ala Asn Asn Gly Gly Val
1 5

<210> SEQ ID NO 79

<211> LENGTH: 8

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 79

Lys Ile Ala Lys Arg Gly Gly Val
1 5

<210> SEQ ID NO 80

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 80

Lys Ile Ala Ser Asn Gly Gly Gly Val
1 5

<210> SEQ ID NO 81

<211> LENGTH: 9

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 81

Lys Ile Ala Ser Asn Ile Gly Gly Val
1 5

<210> SEQ ID NO 82
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 82

Lys Ile Ala Ser His Asp Gly Gly Val
1 5

<210> SEQ ID NO 83
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 83

Lys Ile Ala Ser Asn Asn Gly Gly Val
1 5

<210> SEQ ID NO 84
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 84

Lys Ile Ala Lys Arg Gly Gly Val
1 5

<210> SEQ ID NO 85
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 85

Lys Ile Ala Lys Asn Gly Gly Gly Val
1 5

<210> SEQ ID NO 86
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 86

Lys Ile Ala Lys Asn Ile Gly Gly Val
1 5

<210> SEQ ID NO 87
<211> LENGTH: 9

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 87

Lys Ile Ala Lys His Asp Gly Gly Val
1 5

<210> SEQ ID NO 88
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 88

Lys Ile Ala Lys Asn Asn Gly Gly Val
1 5

<210> SEQ ID NO 89
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 89

Lys Ile Ala Lys Arg Gly Gly Val
1 5

<210> SEQ ID NO 90
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 90

Lys Ile Ala Ser Asn Gly Gly Gly Lys
1 5

<210> SEQ ID NO 91
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 91

Lys Ile Ala Ser Asn Ile Gly Gly Lys
1 5

<210> SEQ ID NO 92
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 92

Lys Ile Ala Ser His Asp Gly Gly Lys
1 5

-continued

<210> SEQ ID NO 93
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 93

Lys Ile Ala Ser Asn Asn Gly Gly Lys
1 5

<210> SEQ ID NO 94
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 94

ctgaccccg caccagtggt ggccatcgcc agcmayggng gcggcaagca ggcgctggag 60
acgggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gc 102

<210> SEQ ID NO 95
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 95

ctgaccccg caccagtggt ggccatcgcc agcmaytcng gcggcaagca ggcgctggag 60
acgggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gc 102

<210> SEQ ID NO 96
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 96

ctgaccccg caccagtggt ggccatcgcc agcmayagy gcggcaagca ggcgctggag 60
acgggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gc 102

<210> SEQ ID NO 97
<211> LENGTH: 102
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 97

ctgaccccg caccagtggt ggccatcgcc agcmayathg gcggcaagca ggcgctggag 60
acgggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gc 102

-continued

<210> SEQ ID NO 98
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 98

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
20 25 30

His Gly

<210> SEQ ID NO 99
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 99

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
20 25 30

His Gly

<210> SEQ ID NO 100
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 100

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys
1 5 10 15

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
20 25 30

His Gly

<210> SEQ ID NO 101
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 101

ctgaccccg caccagtggt ggccatgcc agcmayggng gcggcaagca ggcgctcgag 60

agc 63

<210> SEQ ID NO 102
<211> LENGTH: 63
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: 39
<223> OTHER INFORMATION: n = A, T, C, or G

<400> SEQUENCE: 102

ctgaccccg c cacaggtggt ggccatcgcc agcmaytcng gcggcaagca ggcgctcgag 60
agc 63

<210> SEQ ID NO 103
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 103

ctgaccccg c cacaggtggt ggccatcgcc agcmayagyg gcggcaagca ggcgctcgag 60
agc 63

<210> SEQ ID NO 104
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 104

ctgaccccg c cacaggtggt ggccatcgcc agcmayathg gcggcaagca ggcgctcgag 60
agc 63

<210> SEQ ID NO 105
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 105

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys
1 5 10 15
Gln Ala Leu Glu Ser
20

<210> SEQ ID NO 106
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 106

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys
1 5 10 15
Gln Ala Leu Glu Ser
20

<210> SEQ ID NO 107

-continued

```

<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic peptide

<400> SEQUENCE: 107

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys
1      5      10      15
Gln Ala Leu Glu Ser
      20

```

1. (canceled)
2. A first transcription activator-like (TAL) effector endonuclease monomer comprising:
 - (i) a FokI endonuclease domain, and
 - (ii) a TAL effector domain comprising a plurality of TAL effector repeat sequences, wherein the plurality of TAL effector repeat sequences of the first TAL effector endonuclease monomer, in combination, bind to a first nucleotide sequence in a target DNA sequence in a cell, wherein the first TAL effector endonuclease monomer is capable of forming a dimer with a second TAL effector endonuclease monomer comprising a FokI endonuclease domain and a TAL effector domain having a plurality of TAL effector repeat sequences that, in combination, bind to a second nucleotide sequence in the target DNA sequence, wherein the dimer is formed between the FokI domain of the first TAL effector endonuclease monomer and the FokI domain of the second TAL effector endonuclease monomer when the TAL effector domain of the first TAL effector endonuclease monomer is bound to the first nucleotide sequence and the TAL effector domain of the second TAL effector endonuclease monomer is bound to the second nucleotide sequence, wherein the first nucleotide sequence and the second nucleotide sequence are different and are separated by a spacer sequence, and wherein the dimer cleaves the target DNA sequence within the cell.
3. The first TAL effector endonuclease monomer of claim 2, wherein the target DNA sequence is in a promoter region.
4. The first TAL effector endonuclease monomer of claim 2, wherein the TAL effector domain comprises 15 or more DNA binding repeats.
5. The first TAL effector endonuclease monomer of claim 2, wherein each DNA binding repeat comprises a repeat variable-diresidue (RVD) that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises one or more of:

HD for recognizing C;
 NG for recognizing T;
 NI for recognizing A;
 NN for recognizing G;
 NS for recognizing A;
 HG for recognizing T;
 IG for recognizing T;
 NK for recognizing G;
 HA for recognizing C;
 ND for recognizing C;
 HI for recognizing C;
 HN for recognizing G; and
 NA for recognizing G.

6. The first TAL effector endonuclease monomer of claim 5, wherein each DNA binding repeat comprises a RVD that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises one or more of:

HA for recognizing C;
 ND for recognizing C;
 HI for recognizing C;
 HN for recognizing G;
 NA for recognizing G; and
 NK for recognizing G;

and one or more of:

HD for recognizing C;
 NG for recognizing T;
 NI for recognizing A;
 NN for recognizing G or A;
 NS for recognizing A or C or G or T;
 HG for recognizing T; and
 IG for recognizing T.

7. The first TAL effector endonuclease monomer of claim 1, wherein the spacer sequence is 12 to 30 nucleotides in length.

8. The first TAL effector endonuclease monomer of claim 1, wherein the spacer sequence is 18 nucleotides in length.

* * * * *