

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2012284365 B2**

(54) Title
Methods of transcription activator like effector assembly

(51) International Patent Classification(s)
C12N 15/66 (2006.01) **C12N 15/31** (2006.01)
C07K 14/24 (2006.01) **C12N 15/62** (2006.01)
C07K 19/00 (2006.01) **C12Q 1/68** (2006.01)
C12N 1/00 (2006.01) **C40B 40/06** (2006.01)
C12N 5/00 (2006.01)

(21) Application No: **2012284365** (22) Date of Filing: **2012.07.12**

(87) WIPO No: **WO13/012674**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/610,212	2012.03.13	US
61/601,409	2012.02.21	US
61/508,366	2011.07.15	US

(43) Publication Date: **2013.01.24**

(44) Accepted Journal Date: **2017.04.20**

(71) Applicant(s)
The General Hospital Corporation

(72) Inventor(s)
Joung, J. Keith;Sander, Jeffry D.

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

(56) Related Art
WO 2011/072246 A2
Li T et al, Nucleic Acids Research, March 2011,39(14):6315-6325
WO 2012/138939 A1
Morbitzer R et al, Nucleic Acids Research, March 2011, 39(13):5790-5799



(51) International Patent Classification:

C12N 15/66 (2006.01) C07K 14/24 (2006.01)
C12N 15/62 (2006.01) C12N 1/00 (2006.01)
C12N 15/31 (2006.01) C12N 5/00 (2006.01)
C12Q 1/68 (2006.01) C40B 40/06 (2006.01)
C07K 19/00 (2006.01)

(21) International Application Number:

PCT/US2012/046451

(22) International Filing Date:

12 July 2012 (12.07.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/508,366 15 July 2011 (15.07.2011) US
61/601,409 21 February 2012 (21.02.2012) US
61/610,212 13 March 2012 (13.03.2012) US

(71) Applicant (for all designated States except US): **THE GENERAL HOSPITAL CORPORATION** [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JOUNG, J., Keith** [US/US]; 1 Magnolia Way, Winchester, MA 01890 (US).
SANDER, Jeffrey, D. [US/US]; 111 Locust St Bldg #3 Apt A1, Woburn, MA 01801 (US).

(74) Agents: **LODOVICE, Ian, J.** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, MN 55440-1022 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

[Continued on next page]

(54) Title: METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY

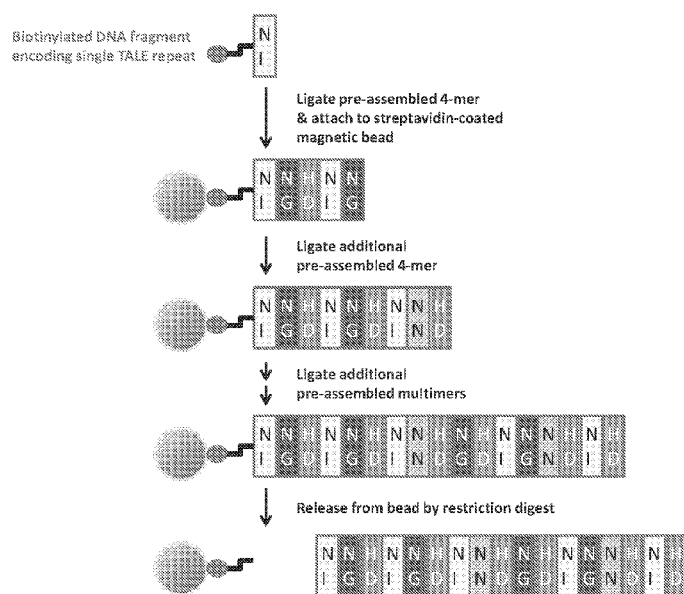


FIG. 1

(57) Abstract: The disclosure describes methods that include providing a first nucleic acid having a sequence encoding a first set comprising one or more transcription activator-like effector (TALE) repeat domains and/or one or more portions of one or more TALE repeat domains; contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; providing a second nucleic acid having a sequence encoding a second set comprising one or more TALE repeat domains and/or one or more portions of one or more TALE repeat domains; contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets.



TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

— *before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY

CLAIM OF PRIORITY

This application claims the benefit of U.S. Provisional Patent Application Serial No. 61/508,366, filed on July 15, 2011, and 61/601,409, filed on February 21, 2012, and 61/610,212, filed on March 13, 2012, and the entire contents of each of the foregoing
5 applications are hereby incorporated by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

This invention was made with government support under grant number DP1 OD006862 awarded by the National Institutes of Health. The government has certain rights in the invention.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on July 11, 2012, is named 2953936W.txt and is 459,673 bytes in size.

TECHNICAL FIELD

This invention relates to methods of producing nucleic acids encoding peptides and polypeptides encoding multiple transcription-like activator effector (TALE) repeat domains and the proteins themselves.

BACKGROUND

20 TALE proteins 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 Romer et al., 2007, Science 318:645).

Specificity for nucleic acid sequences depends on an effector-variable number of imperfect, typically ~33-35 amino acid repeats (Schornack et al., 2006, J. Plant Physiol. 163:256). Each repeat binds to one nucleotide in the target sequence, and the specificity of each repeat for its nucleotide is largely context-independent, allowing for the development of custom sequence-specific TALE proteins (Moscou et al., 2009, Science 326:1501; Boch et al., 2009, Science 326:1509-1512).

SUMMARY

This application is based, at least in part, on the development of rapid, simple, and easily automatable methods for assembling nucleic acids encoding custom TALE repeat array proteins.

Accordingly, this disclosure features a process that includes: (a) providing a first nucleic acid having a sequence encoding a first set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) transcription activator-like effector (TALE) repeat domains and/or one or more portions of one or more TALE repeat domains; (b) contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end; (c) providing a second nucleic acid having a sequence encoding a second set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (d) contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and (e) ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid, wherein the first ligated nucleic acid is linked to a solid support, and wherein the first ligated nucleic acid encodes a polypeptide comprising said first and second sets.

In some embodiments, the methods include linking the first nucleic acid to a solid support prior to (b) contacting the first nucleic acid with the first enzyme or prior to (e) ligating the first and second nucleic acids. In some embodiments, the methods include linking the first ligated nucleic acid to a solid support.

5 In some embodiments, the first set is N-terminal to the second set in the polypeptide. In some embodiments, the second set is N-terminal to the first set in the polypeptide.

In some embodiments, the first and second enzymes are a first and second restriction endonuclease, wherein the first restriction endonuclease cleaves at a site within
10 the first nucleic acid and creates a first cut end, and the second restriction endonuclease cleaves at a site within the second nucleic acid and creates a second cut end, and wherein the first and second ligatable ends are the first and second cut ends. When restriction endonucleases are used, the first ligated nucleic acid cannot include a restriction site recognized by the first restriction endonuclease.

15 The process can further include: (f) contacting the first ligated nucleic acid with a third enzyme, wherein the third enzyme creates a third ligatable end; (g) providing a third nucleic acid comprising a sequence encoding a third set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four,
20 two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (h) contacting the third nucleic acid with a fourth enzyme, wherein the fourth enzyme creates a fourth ligatable end, and wherein the third and fourth ligatable ends are compatible; and (i) ligating the first ligated and third nucleic acids through the
25 third and fourth ligatable ends to produce a second ligated nucleic acid linked to the solid support, wherein the second ligated nucleic acid encodes a polypeptide comprising said first, second, and third sets.

In some embodiments, the third and fourth enzymes are a third and fourth restriction endonuclease, wherein the third restriction endonuclease cleaves at a site
30 within the first ligated nucleic acid and creates a third cut end, and the fourth restriction

endonuclease cleaves at a site within the third nucleic acid and creates a fourth cut end, and wherein the third and fourth ligatable ends are the third and fourth cut ends.

In some embodiments, the ligated nucleic acid does not include a restriction site recognized by the first endonuclease, and the first and third restriction endonucleases are the same. In some embodiments, the second and fourth restriction endonucleases are the same.

The process can further include: (j) contacting the second ligated nucleic acid with a fifth enzyme, wherein the fifth enzyme creates a fifth ligatable end; (k) providing a fourth nucleic acid having a sequence encoding a fourth set comprising one or more (e.g., two or more, three or more, four or more, five or more, six or more, one to six, two to six, three to six, four to six, five or six, one two to five, three to five, four or five, one to four, two to four, three or four, one to three, two or three, one or two, one, two, three, four, five, or six) TALE repeat domains and/or one or more portions of one or more TALE repeat domains; (l) contacting the fourth nucleic acid with a sixth enzyme, wherein the sixth enzyme creates a sixth ligatable end, and wherein the fifth and sixth ligatable ends are compatible; and (m) ligating the second ligated and fourth nucleic acids through the fifth and sixth ligatable ends to produce a third ligated nucleic acid linked to the solid support, wherein the third ligated nucleic acid encodes a polypeptide comprising said first, second, third, and fourth sets. One of ordinary skill would recognize that the process can be repeated with similar additional steps. Such methods are included within this disclosure.

In some embodiments, the fifth and sixth enzymes are a fifth and sixth restriction endonuclease, wherein the fifth restriction endonuclease cleaves at a site within the second ligated nucleic acid and creates a fifth cut end, and the sixth restriction endonuclease cleaves at a site within the fourth nucleic acid and creates a sixth cut end, and wherein the fifth and sixth ligatable ends are the fifth and sixth cut ends.

In some embodiments, the second ligated nucleic acid does not include a restriction site recognized by the first endonuclease, and the first, third, and fifth restriction endonucleases are the same.

In some embodiments, the second, fourth, and sixth restriction endonucleases are the same.

In some embodiments, the solid support and linked nucleic acid are isolated, e.g., following any of the above steps (a) – (m).

5 In some embodiments, the second, third, or fourth set comprises one to four TALE repeat domains.

In some embodiments, the ligatable ends include an overhang of 1-10 nucleotides. In some embodiments, the ligatable ends are blunt ends. In some embodiments, an overhang can be generated using an exonuclease and polymerase in the presence of one
10 or more nucleotides.

In some embodiments, an enzyme or restriction endonuclease used in the above processes is a type IIS restriction endonuclease.

The processes can further comprise unlinking a ligated nucleic acid from the solid support and inserting the ligated nucleic acid (or a processed derivative thereof
15 comprising the TALE repeat array coding sequences) into a vector, e.g., an expression vector. The expression vector can include a sequence encoding an effector domain (e.g., a nuclease domain) configured to create a sequence encoding a fusion protein of the polypeptide and the effector domain. The expression vector can be inserted into a cell to affect the cell directly or for expression of the polypeptide or fusion protein. When the
20 polypeptide or fusion protein is to be expressed, the processes can further include expressing and purifying the polypeptide or fusion protein.

In another aspect, this disclosure features TALE proteins that bind to a target nucleotide sequence (e.g., a “half site”) disclosed herein (e.g., in Table 6 or 7), TALE nucleases that include the TALE proteins, pairs of TALE proteins (e.g., TALENs) that
25 bind to the target sites disclosed herein (e.g., in Table 6 or 7), and nucleic acids that encode any of the above. In some embodiments, the TALE proteins, TALE nucleases, and pairs of TALE proteins (e.g., TALENs) are those disclosed in Example 7. The nucleic acids encoding the TALE proteins, TALE nucleases, and pairs of TALE proteins (e.g., TALENs) can be those disclosed in Example 7 or other sequences that encode the
30 proteins disclosed in Example 7. The disclosure also includes vectors and cells that

include the nucleic acids encoding the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) disclosed herein and methods of expressing the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) that include culturing the cells. The methods of expressing the TALE proteins, TALE nucleases, or
5 pairs of TALE proteins (e.g., TALENs) can also include isolating the TALE proteins, TALE nucleases, or pairs of TALE proteins (e.g., TALENs) from the cell culture.

In another aspect, the invention features a set, archive, or library of nucleic acids (e.g., plasmids) that include sequences encoding one or more TALE domains. In some embodiments, the set, archive, or library includes sequences encoding one, two, three,
10 and/or four (or more than four (e.g., five, six, or more)) TALE repeat domains. In some embodiments, the set, library, or archive of nucleic acids includes sequences encoding TALE repeat domains that bind to nucleotide sequences having one, two, three, four (or more than four (e.g., five, six, or more)) nucleotides. In some embodiments, the set, library, or archive includes restriction sites (e.g., sites for type IIS restriction
15 endonucleases) surrounding the sequences encoding the TALE repeat domains.

The methods described herein provide several advantages, including avoiding extensive PCR amplification of the TALE repeats, thereby avoiding the introduction of mutations from PCR errors. Further, TALE repeat arrays of any desired length can be constructed, and the methods can be easily multiplexed and/or automated.

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 belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications,
20 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.

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
30

advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic depiction of an exemplary method of assembling a nucleic acid encoding a TALE protein.

FIG. 2 is a schematic depiction of exemplary archives of nucleic acids encoding single (one-mer), two-mer, three-mer, and four-mer TALE repeat domains.

FIG. 3 depicts the sequence of the pUC57- Δ BsaI plasmid. This plasmid is identical to plasmid pUC57 except for mutation of a single base (in bold, underlined and lowercase) that destroys a BsaI restriction site.

FIG. 4A depicts the polypeptide sequences of exemplary TALE repeats of type α/ϵ , β , γ , and δ . Polymorphic residues characteristic of each type are indicated in bold and italic. The hypervariable triplet SNI for binding to A is indicated in underscore.

FIG. 4B depicts the polynucleotide sequences of the exemplary TALE repeats of FIG. 4A.

FIGS. 5A-5B depict the common sequence of expression plasmids pJDS70, pJDS71, pJDS74, pJDS76, and pJDS78. The region of the variable sequences is depicted as XXXXXXXXXX (underlined and bold).

FIG. 6 is a schematic diagram of the enhanced green fluorescent protein (eGFP) gene and the location of the binding sites for synthetic TALE proteins described herein.

FIG. 7 is a bar graph depicting the % of TALE nuclease-modified, eGFP-negative cells at 2 and 5 days following transfection with plasmids encoding TALE nucleases designed to bind and cleave the eGFP reporter gene.

FIG. 8 is a depiction of the sequences of insertion-deletion mutants of eGFP induced by TALE nucleases. Deleted bases are indicated by dashes and inserted bases indicated by double underlining; the TALEN target half-sites are single underlined. The net number of bases inserted or deleted is shown to the right.

FIG. 9 is a depiction of an electrophoresis gel of assembled DNA fragments encoding 17-mer TALE array preparations.

FIG. 10 is a depiction of an electrophoresis gel of 16-mer TALE array preparations.

FIGs. 11A-11B depict the nucleotide (11A) and polypeptide (11B) sequence of engineered DR-TALE-0003.

5 FIGs. 12A-12B depict the nucleotide (12A) and polypeptide (12B) sequence of engineered DR-TALE-0006.

FIGs. 13A-13B depict the nucleotide (13A) and polypeptide (13B) sequence of engineered DR-TALE-0005.

10 FIGs. 14A-14B depict the nucleotide (14A) and polypeptide (14B) sequence of engineered DR-TALE-0010.

FIGs. 15A-15B depict the nucleotide (15A) and polypeptide (15B) sequence of engineered DR-TALE-0023.

FIGs. 16A-16B depict the nucleotide (16A) and polypeptide (16B) sequence of engineered DR-TALE-0025.

15 FIGs. 17A-17B depict the nucleotide (17A) and polypeptide (17B) sequence of engineered DR-TALE-0020.

FIGs. 18A-18B depict the nucleotide (18A) and polypeptide (18B) sequence of engineered DR-TALE-0022.

20 FIG. 19A is a bar graph depicting activities of 48 TALEN pairs and four ZFN pairs in the EGFP gene-disruption assay. Percentages of EGFP-negative cells as measured 2 and 5 days following transfection of U2OS cells bearing a chromosomally integrated EGFP reporter gene with nuclease-encoding plasmids are shown. Mean percent disruption of EGFP and standard error of the mean from three independent transfections are shown.

25 FIG. 19B is a bar graph depicting mean EGFP-disruption activities from FIG. 19A, grouped by length of the TALENs.

30 FIG. 20A is a graph depicting the ratio of mean percent EGFP disruption values from day 2 to day 5. Ratios were calculated for groups of each length TALEN using the data from FIG. 19B. Values greater than 1 indicate a decrease in the average of EGFP-disrupted cells at day 5 relative to day 2.

FIG. 20B is a graph depicting the ratio of mean tdTomato-positive cells from day 2 to day 5 grouped by various lengths of TALENs. tdTomato-encoding control plasmids were transfected together with nuclease-encoding plasmids on day 0.

FIGs. 21A-E depict DNA sequences and frequencies of assembled TALEN-induced mutations at endogenous human genes. For each endogenous gene target, the wild-type (WT) sequence is shown at the top with the TALEN target half-sites underlined and the translation start codon of the gene (ATG) indicated by a box. Deletions are indicated by dashes and insertions by lowercase letters and double underlining. The sizes of the insertions (+) or deletions (Δ) are indicated to the right of each mutated site. The number of times that each mutant was isolated is shown in parentheses. Mutation frequencies are calculated as the number of mutants identified divided by the total number of sequences analyzed. Note that for several of the genes, we also identified larger deletions that extend beyond the sequences of the TALEN target sites.

FIG. 22 is a schematic depiction of an exemplary method of assembling a nucleic acid encoding a TALE protein containing TALE repeat domains or portions of TALE repeat domains.

DETAILED DESCRIPTION

The methods described herein can be used to assemble engineered proteins containing TALE repeat domains for binding to specific sequences of interest. Assembling long arrays (e.g., 12 or more) of TALE repeat domain repeats can be challenging because the repeats differ only at a small number of amino acids within their highly conserved ~33-35 amino acid consensus sequence. PCR assembly can lead to the introduction of unwanted mutations. Hierarchical assembly methods that involve one or more passages of intermediate plasmid constructs in *E. coli* can also be problematic because the highly repetitive nature of these constructs can make them unstable and prone to recombination and because the need to passage these intermediate constructs makes these approaches difficult to automate.

TAL Effectors

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. Specificity depends on an effector-variable number of imperfect, typically ~33-35 amino acid repeats. Polymorphisms are present primarily at repeat positions 12 and 13, which are referred to herein as the “repeat variable-diresidue” (RVD). 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. In some embodiments, the polymorphic region that grants nucleotide specificity may be expressed as a triresidue or triplet e.g., encompassing residues 11, 12, and 13.

Each DNA binding repeat can include an 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, but is not limited to, one or more of the following: 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.

TALE proteins are 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 are 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.

Assembly Methods

An example of the methods described herein of assembling a TALE repeat domain array is shown in FIG. 1 and includes the following steps: (1) provision a single biotinylated PCR product encoding one single N-terminal TALE repeat domain (a one-mer) with a linker suitable for attachment to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support); (2) creation of an overhang at the 3' end of the one-mer DNA (e.g., using a Type IIS restriction enzyme); (3) ligation of a second fragment containing four TALE repeat domain (i.e., a pre-assembled four-mer), creating a five-mer; (4) attachment of the five-mer to the solid support; (5) ligation of additional pre-assembled TALE repeat domains to create a long array, e.g., a piece or pieces of DNA encoding one, two, three, or four TALE repeat domains depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product). The final fragment can then be prepared for ligation to an appropriate expression plasmid.

Alternatively, the method can proceed as follows: (1) attachment of a single biotinylated PCR product encoding one single N-terminal TALE repeat domains to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports such as the streptavidin-coated wells of a multi-well plate can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support), (2) creation of an overhang at the 3' end of the anchored DNA (e.g., using a Type IIS restriction enzyme), (3) ligation of a second fragment containing four TALE repeat domain, (4) additional cycles of steps (2) and (3) to create a long array, (5) in the final cycle performing ligation of a piece of DNA encoding one, two, three, or four TALE repeat domains depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product).

Another example of a method of assembling a TALE repeat domain array based on the methods described herein is shown in FIG. 22 and includes the following steps: (1) provision a single biotinylated PCR product encoding a portion of one single N-terminal TALE repeat domain (a partial one-mer) with a linker suitable for attachment to a solid support (in the example shown here, a magnetic streptavidin coated bead is used but other solid supports can also be utilized as well as other ways of tethering the initial DNA fragment to the solid support); (2) creation of an overhang at the 3' end of the partial one-mer DNA (e.g., using a Type IIS restriction enzyme); (3) ligation of a second fragment containing consisting of two partial and three full TALE repeats; (4) attachment of the second fragment to the solid support; (5) ligation of additional pre-assembled TALE repeat domains or portions of TALE repeat domains to create a long array, e.g., a piece or pieces of DNA encoding one, two, three, or four TALE repeat domains (or portions of TALE repeat domains) depending upon the length of the desired final array, and (6) release of the extended DNA encoding the TALE repeats from the solid support (e.g., by using a Type IIS restriction enzyme whose site is built in at the 5' end of the initial biotinylated DNA product). The final fragment can then be prepared for ligation to an appropriate expression plasmid.

The initial nucleic acid encoding one or more TALE repeat domains (or portions) is linked to a solid support. The initial nucleic acid can be prepared by any means (e.g., chemical synthesis, PCR, or cleavage from a plasmid). Additionally, the nucleic acid can be linked to the solid support by any means, e.g., covalently or noncovalently.

In some embodiments, the nucleic acid is linked noncovalently by using a nucleic acid modified with one member of a binding pair and incorporating the other member of the binding pair on the solid support. A member of a binding pair is meant to be one of a first and a second moiety, wherein said first and said second moiety have a specific binding affinity for each other. Suitable binding pairs for use in the invention include, but are not limited to, antigens/antibodies (for example, digoxigenin/anti-digoxigenin, dinitrophenyl (DNP)/anti-DNP, dansyl-X/anti-dansyl, Fluorescein/anti-fluorescein, lucifer yellow/anti-lucifer yellow, peptide/anti-peptide, ligand/receptor and rhodamine/anti-rhodamine), biotin/avidin (or biotin/streptavidin) and calmodulin binding

protein (CBP)/calmodulin. Other suitable binding pairs include polypeptides such as the FLAG-peptide (Hopp et al., 1988, BioTechnology, 6:1204-10); the KT3 epitope peptide (Martin et al., Science 255:192-194 (1992)); tubulin epitope peptide (Skinner et al., J. Biol. Chem. 266:15163-66 (1991)); and the T7 gene 10 protein peptide tag (Lutz-Freyer et al., Proc. Natl. Acad. Sci. USA, 87:6393-97 (1990)) and the antibodies each thereto.

In some embodiments, the individual nucleic acids encoding one or more TALE repeat domains are present in an archive or library of plasmids (see FIG. 2). Although nucleic acids encoding one to four TALE repeat domains are shown, the library of plasmids can contain nucleic acids encoding more than four (e.g., five, six, or more) TALE repeat domains. Alternatively, as shown Figure 22, the nucleic acids encoding parts or portions of one or more TALE repeat domains can also be joined together to create final DNA fragments encoding the desired full-length arrays of TALE repeat domains. Numerous TALE repeat domain sequences with binding specificity for specific nucleotides or sets of nucleotides are known in the art, and one of ordinary skill can design and prepare a library of plasmids based on these known sequences and the disclosures herein.

As used herein, a solid support refers to any solid or semisolid or insoluble support to which the nucleic acid can be linked. Such materials include any materials that are used as supports for chemical and biological molecule syntheses and analyses, such as, but not limited to: polystyrene, polycarbonate, polypropylene, nylon, glass, dextran, chitin, sand, pumice, agarose, polysaccharides, dendrimers, buckyballs, polyacryl-amide, silicon, rubber, and other materials used as supports for solid phase syntheses, affinity separations and purifications, hybridization reactions, immunoassays and other such applications. The solid support can be particulate or can be in the form of a continuous surface, such as a microtiter dish or well, a glass slide, a silicon chip, a nitrocellulose sheet, nylon mesh, or other such materials. When particulate, typically the particles have at least one dimension in the 5-10 mm range or smaller. Such particles, referred collectively herein as "beads," are often, but not necessarily, spherical. Such reference, however, does not constrain the geometry of the matrix, which can be any

shape, including random shapes, needles, fibers, and elongated. Roughly spherical “beads,” particularly microspheres that can be used in the liquid phase, also are contemplated. The “beads” can include additional components, such as magnetic or paramagnetic particles (see, e.g., Dynabeads (Dynal, Oslo, Norway)) for separation using magnets, as long as the additional components do not interfere with the methods described herein.

The ligatable ends can be produced by cutting with a restriction endonuclease (e.g., a type II or type IIS restriction endonuclease) or by “chewing back” the end using an enzyme (or enzymes) with exonuclease and polymerase activities in the presence of one or more nucleotides (see, Aslanidis et al., 1990, Nucl. Acids Res., 18:6069-74). Suitable enzymes are known to those of ordinary skill in the art. When restriction endonucleases are used, the nucleic acids can be designed to include restriction sites for the enzymes at suitable locations.

Following a ligation reaction, any unligated ends with 5' or 3' overhangs can be “blunted” by use of a polymerase, e.g., a DNA polymerase with both 3'→5' exonuclease activity and 5'→3' polymerase activity. This blunting step can reduce the appearance of undesired or partial assembly products. Alternatively, these ends can be capped using either a “hairpin” oligo bearing a compatible overhang (Briggs et al., 2012, Nucleic Acids Res, PMID: 22740649) or by short double-stranded DNAs bearing a compatible overhang on one end and a blunt end on the other.

To prepare the ligated nucleic acid for further downstream processing, it can be useful to select nucleic acids of the expected size, to reduce the presence of minor products created by incomplete ligations. Methods of selecting nucleic acids by size are known in the art, and include gel electrophoresis (e.g., slab gel electrophoresis or capillary gel electrophoresis (see, e.g., Caruso et al., 2003, Electrophoresis, 24:1-2:78-85)), liquid chromatography (e.g., size exclusion chromatography or reverse phase chromatography (see, e.g., Huber et al., 1995, Anal. Chem., 67:578-585)), and lab-on-a-chip systems (e.g., LabChip® XT system, Caliper Life Sciences, Hopkinton, MA). In some embodiments, a size exclusion step can be performed using an automated system,

e.g., an automated gel electrophoresis system (e.g., a Pippin Prep™ automated DNA size selection system, Sage Science, Beverly, MA).

Automation

5 The methods disclosed herein can be performed manually or implemented in laboratory automation hardware (e.g., SciClone G3 Liquid Handling Workstation, Caliper Life Sciences, Hopkinton, MA) controlled by a compatible software package (e.g., Maestro™ liquid handling software) programmed according to the new methods described herein or a new software package designed and implemented to carry out the
10 specific method steps described herein. When performed by laboratory automation hardware, the methods can be implemented by computer programs using standard programming techniques following the method steps described herein.

 Examples of automated laboratory system robots include the Sciclone™ G3 liquid handling workstation (Caliper Life Sciences, Hopkinton, MA), Biomek® FX liquid
15 handling system (Beckman-Coulter, Fullerton, California), TekBench™ automated liquid handling platform (TekCel, Hopkinton, Massachusetts), and Freedom EVO® automation platform (Tecan Trading AG, Switzerland).

 The programs can be designed to execute on a programmable computer including at least one processor, at least one data storage system (including volatile and non-
20 volatile memory and/or storage elements, e.g., RAM and ROM), at least one communications port that provides access for devices such as a computer keyboard, telephone, or a wireless, hand-held device, such as a PDA, and optionally at least one output device, such as a monitor, printer, or website. The central computer also includes a clock and a communications port that provides control of the lab automation hardware.
25 These are all implemented using known techniques, software, and devices. The system also includes a database that includes data, e.g., data describing the procedure of one or more method steps described herein.

 Program code is applied to data input by a user (e.g., location of samples to be processed, timing and frequency of manipulations, amounts of liquid dispensed or
30 aspirated, transfer of samples from one location in the system to another) and data in the

database, to perform the functions described herein. The system can also generate inquiries and provide messages to the user. The output information is applied to instruments, e.g., robots, that manipulate, heat, agitate, etc. the vessels that contain the reactants as described herein. In addition, the system can include one or more output
5 devices such as a telephone, printer, or a monitor, or a web page on a computer monitor with access to a website to provide to the user information regarding the synthesis and/or its progress.

Each program embodying the new methods is preferably implemented in a high level procedural or object-oriented programming language to communicate with a
10 computer system. However, the programs can also be implemented in assembly or machine language if desired. In any case, the language can be a compiled or interpreted language.

Each such computer program is preferably stored on a storage medium or device (e.g., RAM, ROM, optical, magnetic) readable by a general or special purpose
15 programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system can also be considered to be implemented as a computer- or machine-readable storage medium (electronic apparatus readable medium), configured with a program, whereby the storage medium so configured causes a computer or machine to operate in a
20 specific and predefined manner to perform the functions described herein.

The new methods can be implemented using various means of data storage. The files can be transferred physically on recordable media or electronically, e.g., by email on a dedicated intranet, or on the Internet. The files can be encrypted using standard encryption software from such companies as RSA Security (Bedford, Massachusetts) and
25 Baltimore[®]. The files can be stored in various formats, e.g., spreadsheets or databases.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; communications networks, including
30 local area networks (LAN), wide area networks (WAN), Internet, Intranet, and Extranet;

electronic appliances such as a personal digital assistants (PDAs), cellular telephones, "smartphones," pagers and the like; and local and distributed processing systems.

As used herein, "stored" refers to a process for encoding information on an electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the sequence information.

A variety of software programs and formats can be used to store method data on an electronic apparatus readable medium. For example, the data and machine instructions can be incorporated in the system of the software provided with the automated system, represented in a word processing text file, formatted in commercially-available software such as WordPerfect® and Microsoft® Word®, or represented in the form of an ASCII file, stored in a database application, such as Microsoft Access®, Microsoft SQL Server®, Sybase®, Oracle®, or the like, as well as in other forms. Any number of data processor structuring formats (e.g., text file or database) can be employed to obtain or create a medium having recorded thereon the relevant data and machine instructions to implement the methods described herein.

By providing information in electronic apparatus readable form, the programmable computer can communicate with and control the lab automation hardware to perform the methods described herein. One skilled in the art can input data in electronic apparatus readable form (or a form that is converted to electronic apparatus readable form) to describe the completion of various method steps by the lab automation hardware.

Polypeptide Expression Systems

In order to use the engineered proteins of the present invention, it is typically necessary to express the engineered proteins from a nucleic acid that encodes them. This can be performed in a variety of ways. For example, the nucleic acid encoding the engineered TALE repeat protein is typically cloned into an intermediate vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors, or

insect vectors, for storage or manipulation of the nucleic acid encoding the engineered TALE protein or production of protein. The nucleic acid encoding the engineered TALE repeat protein is also typically cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression of a cloned gene or nucleic acid, the engineered TALE repeat protein is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook et al., *Molecular Cloning, A Laboratory Manual* (3d ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 2010). Bacterial expression systems for expressing the engineered TALE repeat protein are available in, e.g., *E. coli*, *Bacillus* sp., and *Salmonella* (Palva et al., 1983, *Gene* 22:229-235). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

The promoter used to direct expression of the engineered TALE repeat protein nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of the engineered TALE repeat protein. In contrast, when the engineered TALE repeat protein is to be administered in vivo for gene regulation, either a constitutive or an inducible promoter can be used, depending on the particular use of the engineered TALE repeat protein. In addition, a preferred promoter for administration of the engineered TALE repeat protein can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (see, e.g., Gossen & Bujard, 1992, *Proc. Natl. Acad. Sci. USA*, 89:5547; Oligino et al., 1998, *Gene Ther.*, 5:491-496; Wang et al., 1997, *Gene Ther.*, 4:432-441; Neering et al., 1996, *Blood*, 88:1147-55; and Rendahl et al., 1998, *Nat. Biotechnol.*, 16:757-761).

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, e.g., to the nucleic acid sequence encoding the TALE repeat protein signals required, e.g., for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette can include, e.g., enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the engineered TALE repeat protein, e.g., expression in plants, animals, bacteria, fungus, protozoa, etc. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. A preferred fusion protein is the maltose binding protein, "MBP." Such fusion proteins can be used for purification of the engineered TALE repeat protein. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with the engineered TALE repeat protein encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

5 Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (see, e.g., Colley et al., 1989, J. Biol. Chem., 264:17619-22; Guide to Protein Purification, in Methods in Enzymology, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard
10 techniques (see, e.g., Morrison, 1977, J. Bacteriol. 132:349-351; Clark-Curtiss & Curtiss, Methods in Enzymology 101:347-362 (Wu et al., eds, 1983).

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells can be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, naked DNA,
15 plasmid vectors, viral vectors, both episomal and integrative, and any of the other well-known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook et al., supra). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the
20 protein of choice.

Characterization of TALE Proteins

Engineered TALE repeat array proteins designed using methods of the present invention can be further characterized to ensure that they have the desired characteristics
25 for their chosen use. For example, TALE repeat array protein can be assayed using a bacterial two-hybrid, bacterial promoter repression, phage-display, or ribosome display system or using an electrophoretic mobility shift assay or "EMSA" (Buratowski & Chodosh, in Current Protocols in Molecular Biology pp. 12.2.1-12.2.7). Equally, any other DNA binding assay known in the art could be used to verify the DNA binding
30 properties of the selected protein.

In one embodiment, a bacterial "two-hybrid" system is used to express and test a TALE repeat protein of the present invention. The bacterial two-hybrid system has an additional advantage, in that the protein expression and the DNA binding "assay" occur within the same cells, thus there is no separate DNA binding assay to set up.

5 Methods for the use of the bacterial two-hybrid system to express and assay DNA binding proteins are described in Joung et al., 2000, Proc. Natl. Acad. Sci. USA, 97:7382, Wright et al., 2006, Nat. Protoc, 1:1637-52; Maeder et al., 2008, Mol. Cell, 31:294-301; Maeder et al., 2009, Nat. Protoc., 4:1471-1501; and US Patent Application No. 2002/0119498, the contents of which are incorporated herein by reference. Briefly, in a
10 bacterial two-hybrid system, the DNA binding protein is expressed in a bacterial strain bearing the sequence of interest upstream of a weak promoter controlling expression of a reporter gene (e.g., histidine 3 (HIS3), the beta-lactamase antibiotic resistance gene, or the beta-galactosidase (lacZ) gene). Expression of the reporter gene occurs in cells in which the DNA binding protein expressed by the cell binds to the target site sequence.
15 Thus, bacterial cells expressing DNA binding proteins that bind to their target site are identified by detection of an activity related to the reporter gene (e.g., growth on selective media, expression of beta-galactosidase).

 In some embodiments, calculations of binding affinity and specificity are also made. This can be done by a variety of methods. The affinity with which the selected
20 TALE repeat array protein binds to the sequence of interest can be measured and quantified in terms of its K_D . Any assay system can be used, as long as it gives an accurate measurement of the actual K_D of the TALE repeat array protein. In one embodiment, the K_D for the binding of a TALE repeat array protein to its target is measured using an EMSA

25 In one embodiment, EMSA is used to determine the K_D for binding of the selected TALE repeat array protein both to the sequence of interest (i.e., the specific K_D) and to non-specific DNA (i.e., the non-specific K_D). Any suitable non-specific or "competitor" double stranded DNA known in the art can be used. In some embodiments, calf thymus DNA or human placental DNA is used. The ratio of the non-specific K_D to the specific
30 KD is the specificity ratio. TALE repeat array proteins that bind with high specificity

have a high specificity ratio. This measurement is very useful in deciding which of a group of selected TALE should be used for a given purpose. For example, use of TALE repeat array protein in vivo requires not only high affinity binding but also high-specificity binding.

5

Construction of Chimeric TALE Proteins

Often, the aim of producing a custom-designed TALE repeat array DNA binding domain is to obtain a TALE repeat array protein that can be used to perform a function. The TALE repeat array DNA binding domain can be used alone, for example to bind to a specific site on a gene and thus block binding of other DNA-binding domains. However, in some embodiments, the TALE repeat array protein will be used in the construction of a chimeric TALE protein containing a TALE repeat array DNA binding domain and an additional domain having some desired specific function (e.g., gene activation) or enzymatic activity i.e., a "functional domain."

Chimeric TALE repeat array proteins designed and produced using the methods described herein can be used to perform any function where it is desired to target, for example, some specific enzymatic activity to a specific DNA sequence, as well as any of the functions already described for other types of synthetic or engineered DNA binding molecules. Engineered TALE repeat array DNA binding domains, can be used in the construction of chimeric proteins useful for the treatment of disease (see, for example, U.S. patent application 2002/0160940, and U.S. Pat. Nos. 6,511,808, 6,013,453 and 6,007,988, and International patent application WO 02/057308), or for otherwise altering the structure or function of a given gene in vivo. The engineered TALE repeat array proteins of the present invention are also useful as research tools, for example, in performing either in vivo or in vitro functional genomics studies (see, for example, U.S. Pat. No. 6,503,717 and U.S. patent application 2002/0164575).

To generate a functional recombinant protein, the engineered TALE repeat array DNA binding domain will typically be fused to at least one "functional" domain. Fusing functional domains to synthetic TALE repeat array proteins to form functional transcription factors involves only routine molecular biology techniques which are

commonly practiced by those of skill in the art, see for example, U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, 6,503,717 and U.S. patent application 2002/0160940).

Functional domains can be associated with the engineered TALE repeat array domain at any suitable position, including the C- or N-terminus of the TALE protein.

5 Suitable "functional" domains for addition to the engineered protein made using the methods of the invention are described in U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, and 6,503,717 and U.S. patent application 2002/0160940.

In one embodiment, the functional domain is a nuclear localization domain which provides for the protein to be translocated to the nucleus. Several nuclear localization
10 sequences (NLS) are known, and any suitable NLS can be used. For example, many NLSs have a plurality of basic amino acids, referred to as a bipartite basic repeats (reviewed in Garcia-Bustos et al, 1991, Biochim. Biophys. Acta, 1071:83-101). An NLS containing bipartite basic repeats can be placed in any portion of chimeric protein and results in the chimeric protein being localized inside the nucleus. It is preferred that a
15 nuclear localization domain is routinely incorporated into the final chimeric protein, as the ultimate functions of the chimeric proteins of the present invention will typically require the proteins to be localized in the nucleus. However, it may not be necessary to add a separate nuclear localization domain in cases where the engineered TALE repeat array domain itself, or another functional domain within the final chimeric protein, has
20 intrinsic nuclear translocation function.

In another embodiment, the functional domain is a transcriptional activation domain such that the chimeric protein can be used to activate transcription of the gene of interest. Any transcriptional activation domain known in the art can be used, such as for
25 example, the VP16 domain from herpes simplex virus (Sadowski et al., 1988, Nature, 335:563-564) or the p65 domain from the cellular transcription factor NF-kappaB (Ruben et al., 1991, Science, 251:1490-93).

In yet another embodiment, the functional domain is a transcriptional repression domain such that the chimeric protein can be used to repress transcription of the gene of interest. Any transcriptional repression domain known in the art can be used, such as for

example, the KRAB (Kruppel-associated box) domain found in many naturally occurring KRAB proteins (Thiesen et al., 1991, Nucleic Acids Res., 19:3996).

In a further embodiment, the functional domain is a DNA modification domain such as a methyltransferase (or methylase) domain, a de-methylation domain, a
5 deaminase domain, a hydroxylase domain, an acetylation domain, or a deacetylation domain. Many such domains are known in the art and any such domain can be used, depending on the desired function of the resultant chimeric protein. For example, it has been shown that a DNA methylation domain can be fused to a TALE repeat array DNA
10 binding protein and used for targeted methylation of a specific DNA sequence (Xu et al., 1997, Nat. Genet., 17:376-378). The state of methylation of a gene affects its expression and regulation, and furthermore, there are several diseases associated with defects in DNA methylation.

In a still further embodiment the functional domain is a chromatin modification domain such as a histone acetylase or histone de-acetylase (or HDAC) domain. Many
15 such domains are known in the art and any such domain can be used, depending on the desired function of the resultant chimeric protein. Histone deacetylases (such as HDAC1 and HDAC2) are involved in gene repression. Therefore, by targeting HDAC activity to a specific gene of interest using an engineered TALE protein, the expression of the gene of interest can be repressed.

20 In an alternative embodiment, the functional domain is a nuclease domain, such as a restriction endonuclease (or restriction enzyme) domain. The DNA cleavage activity of a nuclease enzyme can be targeted to a specific target sequence by fusing it to an appropriate engineered TALE repeat array DNA binding domain. In this way, sequence specific chimeric restriction enzyme can be produced. Several nuclease domains are
25 known in the art and any suitable nuclease domain can be used. For example, an endonuclease domain of a type IIS restriction endonuclease (e.g., FokI) can be used, as taught by Kim et al., 1996, Proc. Natl. Acad. Sci. USA, 6:1156-60). In some embodiments, the endonuclease is an engineered FokI variant as described in US 2008/0131962. Such chimeric endonucleases can be used in any situation where
30 cleavage of a specific DNA sequence is desired, such as in laboratory procedures for the

construction of recombinant DNA molecules, or in producing double-stranded DNA breaks in genomic DNA in order to promote homologous recombination (Kim et al., 1996, Proc. Natl. Acad. Sci. USA, 6:1156-60; Bibikova et al., 2001, Mol. Cell. Biol., 21:289-297; Porteus & Baltimore, 2003, Science, 300:763; Miller et al., 2011, Nat. Biotechnol., 29:143-148; Cermak et al., 2011, Nucl. Acids Res., 39:e82). Repair of TALE nuclease-induced double-strand breaks (DSB) by error-prone non-homologous end-joining leads to efficient introduction of insertion or deletion mutations at the site of the DSB (Miller et al., 2011, Nat. Biotechnol., 29:143-148; Cermak et al., 2011, Nucl. Acids Res., 39:e82). Alternatively, repair of a DSB by homology-directed repair with an exogenously introduced "donor template" can lead to highly efficient introduction of precise base alterations or insertions at the break site (Bibikova et al., 2003, Science, 300:764; Urnov et al., 2005, Nature, 435:646-651; Porteus et al., 2003, Science, 300:763; Miller et al., 2011, Nat. Biotechnol., 29:143-148).

In some embodiments, the functional domain is an integrase domain, such that the chimeric protein can be used to insert exogenous DNA at a specific location in, for example, the human genome.

Other suitable functional domains include silencer domains, nuclear hormone receptors, resolvase domains oncogene transcription factors (e.g., myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.), kinases, phosphatases, and any other proteins that modify the structure of DNA and/or the expression of genes. Suitable kinase domains, from kinases involved in transcription regulation are reviewed in Davis, 1995, Mol. Reprod. Dev., 42:459-67. Suitable phosphatase domains are reviewed in, for example, Schonthal & Semin, 1995, Cancer Biol. 6:239-48.

Fusions of TALE repeat arrays to functional domains can be performed by standard recombinant DNA techniques well known to those skilled in the art, and as are described in, for example, basic laboratory texts such as Sambrook et al., Molecular Cloning; A Laboratory Manual 3d ed. (2001), and in U.S. Pat. Nos. 6,511,808, 6,013,453, 6,007,988, and 6,503,717 and U.S. patent application 2002/0160940.

In some embodiments, two or more engineered TALE repeat array proteins are linked together to produce the final DNA binding domain. The linkage of two or more

engineered proteins can be performed by covalent or non-covalent means. In the case of covalent linkage, engineered proteins can be covalently linked together using an amino acid linker (see, for example, U.S. patent application 2002/0160940, and International applications WO 02/099084 and WO 01/53480). This linker can be any string of amino acids desired. In one embodiment the linker is a canonical TGEKP linker. Whatever linkers are used, standard recombinant DNA techniques (such as described in, for example, Sambrook et al., Molecular Cloning; A Laboratory Manual 3d ed. (2001)) can be used to produce such linked proteins.

In embodiments where the engineered proteins are used in the generation of chimeric endonuclease, the chimeric protein can possess a dimerization domain as such endonucleases are believed to function as dimers. Any suitable dimerization domain can be used. In one embodiment the endonuclease domain itself possesses dimerization activity. For example, the nuclease domain of FokI which has intrinsic dimerization activity can be used (Kim et al., 1996, Proc. Natl. Acad. Sci., 93:1156-60).

Assays for Determining Regulation of Gene Expression by Engineered Proteins

A variety of assays can be used to determine the level of gene expression regulation by the engineered TALE repeat proteins, see for example U.S. Pat. No. 6,453,242. The activity of a particular engineered TALE repeat protein can be assessed using a variety of in vitro and in vivo assays, by measuring, e.g., protein or mRNA levels, product levels, enzyme activity, tumor growth; transcriptional activation or repression of a reporter gene; second messenger levels (e.g., cGMP, cAMP, IP3, DAG, Ca²⁺); cytokine and hormone production levels; and neovascularization, using, e.g., immunoassays (e.g., ELISA and immunohistochemical assays with antibodies), hybridization assays (e.g., RNase protection, northern, in situ hybridization, oligonucleotide array studies), colorimetric assays, amplification assays, enzyme activity assays, tumor growth assays, phenotypic assays, and the like.

TALE proteins can be first tested for activity in vitro using cultured cells; e.g., 293 cells, CHO cells, VERO cells, BHK cells, HeLa cells, COS cells, and the like. In some embodiments, human cells are used. The engineered TALE repeat array protein is

often first tested using a transient expression system with a reporter gene, and then regulation of the target endogenous gene is tested in cells and in animals, both in vivo and ex vivo. The engineered TALE repeat array protein can be recombinantly expressed in a cell, recombinantly expressed in cells transplanted into an animal, or recombinantly
5 expressed in a transgenic animal, as well as administered as a protein to an animal or cell using delivery vehicles described below. The cells can be immobilized, be in solution, be injected into an animal, or be naturally occurring in a transgenic or non-transgenic animal.

Modulation of gene expression is tested using one of the in vitro or in vivo assays
10 described herein. Samples or assays are treated with the engineered TALE repeat array protein and compared to un-treated control samples, to examine the extent of modulation. For regulation of endogenous gene expression, the TALE repeat array protein ideally has a K_D of 200 nM or less, more preferably 100 nM or less, more preferably 50 nM, most preferably 25 nM or less. The effects of the engineered TALE repeat array protein can be
15 measured by examining any of the parameters described above. Any suitable gene expression, phenotypic, or physiological change can be used to assess the influence of the engineered TALE repeat array protein. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as tumor growth, neovascularization, hormone release, transcriptional changes to both
20 known and uncharacterized genetic markers (e.g., northern blots or oligonucleotide array studies), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as cGMP.

Preferred assays for regulation of endogenous gene expression can be performed in vitro. In one in vitro assay format, the engineered TALE repeat array protein
25 regulation of endogenous gene expression in cultured cells is measured by examining protein production using an ELISA assay. The test sample is compared to control cells treated with an empty vector or an unrelated TALE repeat array protein that is targeted to another gene.

In another embodiment, regulation of endogenous gene expression is determined
30 in vitro by measuring the level of target gene mRNA expression. The level of gene

expression is measured using amplification, e.g., using RT-PCR, LCR, or hybridization assays, e.g., northern hybridization, RNase protection, dot blotting. RNase protection is used in one embodiment. The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, a reporter gene system can be devised using the target gene promoter operably linked to a reporter gene such as luciferase, green fluorescent protein, CAT, or beta-galactosidase. The reporter construct is typically co-transfected into a cultured cell. After treatment with the TALE repeat array protein, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

Another example of an assay format useful for monitoring regulation of endogenous gene expression is performed in vivo. This assay is particularly useful for examining TALE repeat array proteins that inhibit expression of tumor promoting genes, genes involved in tumor support, such as neovascularization (e.g., VEGF), or that activate tumor suppressor genes such as p53. In this assay, cultured tumor cells expressing the engineered TALE protein are injected subcutaneously into an immune compromised mouse such as an athymic mouse, an irradiated mouse, or a SCID mouse. After a suitable length of time, preferably 4-8 weeks, tumor growth is measured, e.g., by volume or by its two largest dimensions, and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth. Alternatively, the extent of tumor neovascularization can also be measured.

Immunoassays using endothelial cell specific antibodies are used to stain for vascularization of the tumor and the number of vessels in the tumor. Tumors that have a statistically significant reduction in the number of vessels (using, e.g., Student's T test) are said to have inhibited neovascularization.

Transgenic and non-transgenic animals can also be used for examining regulation of endogenous gene expression in vivo. Transgenic animals can express the engineered TALE repeat array protein. Alternatively, animals that transiently express the engineered TALE repeat array protein, or to which the engineered TALE repeat array protein has

been administered in a delivery vehicle, can be used. Regulation of endogenous gene expression is tested using any one of the assays described herein.

Use of Engineered TALE Repeat-Containing Proteins in Gene Therapy

5 The engineered proteins of the present invention can be used to regulate gene expression or alter gene sequence in gene therapy applications in the same. Similar methods have been described for synthetic zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,013,453, U.S. Pat. No. 6,007,988, U.S. Pat. No. 6,503,717, U.S. patent application 2002/0164575, and U.S. patent application
10 2002/0160940.

 Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding the engineered TALE repeat array protein into mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding engineered TALE repeat array proteins to cells in vitro. Preferably, the nucleic
15 acids encoding the engineered TALE repeat array proteins are administered for in vivo or ex vivo gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene
20 therapy procedures, see Anderson, 1992, Science, 256:808-813; Nabel & Felgner, 1993, TIBTECH, 11:211-217; Mitani & Caskey, 1993, TIBTECH, 11:162-166; Dillon, 1993, TIBTECH, 11:167-175; Miller, 1992, Nature, 357:455-460; Van Brunt, 1988, Biotechnology, 6:1149-54; Vigne, 1995, Restorat. Neurol. Neurosci., 8:35-36; Kremer & Perricaudet, 1995, Br. Med. Bull., 51:31-44; Haddada et al., in Current Topics in
25 Microbiology and Immunology Doerfler and Bohm (eds) (1995); and Yu et al., 1994, Gene Ther., 1:13-26.

 Methods of non-viral delivery of nucleic acids encoding the engineered TALE repeat array proteins include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA or RNA,
30 artificial virions, and agent-enhanced uptake of DNA or RNA. Lipofection is described

in e.g., U.S. Pat. No. 5,049,386, No. 4,946,787; and No.4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024. Delivery can be to cells (ex vivo administration) or target tissues (in vivo administration).

The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, 1995, Science, 270:404-410; Blaese et al., 1995, Cancer Gene Ther., 2:291-297; Behr et al., 1994, Bioconjugate Chem. 5:382-389; Remy et al., 1994, Bioconjugate Chem., 5:647-654; Gao et al., Gene Ther., 2:710-722; Ahmad et al., 1992, Cancer Res., 52:4817-20; U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding the engineered TALE repeat array proteins takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro and the modified cells are administered to patients (ex vivo). Conventional viral based systems for the delivery of TALE repeat array proteins could include retroviral, lentivirus, adenoviral, adeno-associated, Sendai, and herpes simplex virus vectors for gene transfer. Viral vectors are currently the most efficient and versatile method of gene transfer in target cells and tissues. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal

repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., 1992, J. Virol., 66:2731-39; Johann et al., 1992, J. Virol., 66:1635-40; Sommerfelt et al., 1990, Virology, 176:58-59; Wilson et al., 1989, J. Virol., 63:2374-78; Miller et al., 1991, J. Virol., 65:2220-24; WO 94/26877).

In applications where transient expression of the engineered TALE repeat array protein is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., 1987, Virology 160:38-47; U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, 1994, Hum. Gene Ther., 5:793-801; Muzyczka, 1994, J. Clin. Invest., 94:1351). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., 1985, Mol. Cell. Biol. 5:3251-60; Tratschin et al., 1984, Mol. Cell. Biol., 4:2072-81; Hermonat & Muzyczka, 1984, Proc. Natl. Acad. Sci. USA, 81:6466-70; and Samulski et al., 1989, J. Virol., 63:3822-28.

In particular, at least six viral vector approaches are currently available for gene transfer in clinical trials, with retroviral vectors by far the most frequently used system. All of these viral vectors utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials (Dunbar et al., 1995, Blood, 85:3048; Kohn et al., 1995, Nat. Med., 1:1017; Malech et al., 1997, Proc. Natl. Acad. Sci. USA, 94:12133-38). PA317/pLASN was the

first therapeutic vector used in a gene therapy trial. (Blaese et al., 1995, Science, 270:475-480). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors (Ellem et al., 1997, Immunol Immunother., 44:10-20; Dranoff et al., 1997, Hum. Gene Ther., 1:111-112).

5 Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. Typically, the vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes
10 of the transduced cell are key features for this vector system (Wagner et al., 1998, Lancet, 351:1702-1703; Kearns et al., 1996, Gene Ther., 9:748-55).

 Replication-deficient recombinant adenoviral vectors (Ad) are predominantly used for colon cancer gene therapy, because they can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered
15 such that a transgene replaces the Ad E1a, E1b, and E3 genes; subsequently the replication defector vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues in vivo, including nondividing, differentiated cells such as those found in the liver, kidney and muscle system tissues. Conventional Ad vectors have a large carrying capacity. An example of
20 the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman et al., 1998, Hum. Gene Ther. 7:1083-89). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker et al., 1996, Infection, 24:15-10; Sterman et al., 1998, Hum. Gene Ther., 9:7 1083-89; Welsh et al., 1995, Hum. Gene Ther., 2:205-218; Alvarez
25 et al., 1997, Hum. Gene Ther. 5:597-613; Topf et al., 1998, Gene Ther., 5:507-513; Sterman et al., 1998, Hum. Gene Ther., 7:1083-89.

 Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and Ψ 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated
30 by producer cell line that packages a nucleic acid vector into a viral particle. The vectors

typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. The missing viral functions are supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only
5 possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper
10 plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. A viral vector is
15 typically modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the viruses outer surface. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han et al., 1995, Proc. Natl. Acad. Sci. USA, 92:9747-51, reported that Moloney murine leukemia virus can be modified to express human heregulin fused to
20 gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, filamentous phage can be engineered to display antibody fragments (e.g., Fab or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although
25 the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

Gene therapy vectors can be delivered in vivo by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal,
30 intramuscular, subdermal, or intracranial infusion) or topical application, as described

below. Alternatively, vectors can be delivered to cells ex vivo, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., *Diabetologia*, 47:499-508, 2004)), followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with nucleic acid (gene or cDNA), encoding the engineered TALE repeat array protein, and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art (see, e.g., Freshney et al., *Culture of Animal Cells, A Manual of Basic Technique* (5th ed. 2005)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

In one embodiment, stem cells (e.g., universal donor hematopoietic stem cells, embryonic stem cells (ES), partially differentiated stem cells, non-pluripotent stem cells, pluripotent stem cells, induced pluripotent stem cells (iPS cells) (see e.g., Sipione et al., *Diabetologia*, 47:499-508, 2004)) are used in ex vivo procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types in vitro, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells in vitro into clinically important immune cell types using cytokines such as GM-CSF, IFN-gamma and TNF-alpha are known (see Inaba et al., 1992, *J. Exp. Med.*, 176:1693-1702).

Stem cells can be isolated for transduction and differentiation using known methods. For example, stem cells can be isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+

(T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (see Inaba et al., 1992, J. Exp. Med., 176:1693-1702).

Vectors (e.g., retroviruses, adenoviruses, liposomes, etc.) containing nucleic acids encoding the engineered TALE repeat array protein can be also administered directly to the organism for transduction of cells in vivo. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route. Alternatively, stable formulations of the engineered TALE repeat array protein can also be administered.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

Delivery Vehicles

An important factor in the administration of polypeptide compounds, such as the engineered TALE repeat array proteins of the present invention, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as engineered TALE repeat array protein across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-

translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (see, e.g., Prochiantz, 1996, *Curr. Opin. Neurobiol.*, 6:629-634).

5 Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (see, e.g., Lin et al., 1995, *J. Biol. Chem.*, 270:14255-58).

Examples of peptide sequences that can be linked to a protein, for facilitating uptake of the protein into cells, include, but are not limited to: peptide fragments of the tat protein of HIV (Endoh et al., 2010, *Methods Mol. Biol.*, 623:271-281; Schmidt et al., 10 2010, *FEBS Lett.*, 584:1806-13; Futaki, 2006, *Biopolymers*, 84:241-249); a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (see Fahraeus et al., 1996, *Curr. Biol.*, 6:84); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi et al., 1994, *J. Biol. Chem.*, 269:10444); the h 15 region of a signal peptide, such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin et al., *supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, 1997, *Cell*, 88:223-233). See also, e.g., Caron et al., 2001, *Mol Ther.*, 3:310-318; Langel, *Cell-Penetrating Peptides: Processes and Applications* (CRC Press, Boca Raton FL 2002); El-Andaloussi et al., 2005, *Curr. Pharm. Des.*, 11:3597-3611; and Deshayes et al., 2005, 20 *Cell. Mol. Life Sci.*, 62:1839-49. Other suitable chemical moieties that provide enhanced cellular uptake can also be chemically linked to TALE repeat array proteins described herein.

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary 25 toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), 30 have been used in attempts to deliver peptides to the cell cytosol as internal or amino-

terminal fusions (Arora et al., 1993, J. Biol. Chem., 268:3334-41; Perelle et al., 1993, Infect. Immun., 61:5147-56; Stenmark et al., 1991, J. Cell Biol., 113:1025-32; Donnelly et al., 1993, Proc. Natl. Acad. Sci. USA, 90:3530-34; Carbonetti et al., 1995, Abstr. Annu. Meet. Am. Soc. Microbiol. 95:295; Sebo et al., 1995, Infect. Immun., 63:3851-57; 5 Klimpel et al., 1992, Proc. Natl. Acad. Sci. USA, 89:10277-81; and Novak et al., 1992, J. Biol. Chem., 267:17186-93).

Such subsequences can be used to translocate engineered TALE repeat array proteins across a cell membrane. The engineered TALE repeat array proteins can be conveniently fused to or derivatized with such sequences. Typically, the translocation 10 sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the engineered TALE repeat array protein and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

The engineered TALE repeat array protein can also be introduced into an animal cell, preferably a mammalian cell, via liposomes and liposome derivatives such as 15 immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., the engineered TALE repeat array protein.

The liposome fuses with the plasma membrane, thereby releasing the compound 20 into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (e.g., the engineered TALE 25 repeat array protein or a nucleic acid encoding the same) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active compound release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be 30 constructed so that they become destabilized when the environment becomes acidic near

the liposome membrane (see, e.g., Proc. Natl. Acad. Sci. USA, 84:7851 (1987); Biochemistry, 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

Such liposomes typically comprise the engineered TALE repeat array protein and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g., Szoka et al., 1980, Annu. Rev. Biophys. Bioeng., 9:467, U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication. No. WO 91/17424, Deamer & Bangham, 1976, Biochim. Biophys. Acta, 443:629-634; Fraley, et al., 1979, Proc. Natl. Acad. Sci. USA, 76:3348-52; Hope et al., 1985, Biochim. Biophys. Acta, 812:55-65; Mayer et al., 1986, Biochim. Biophys. Acta, 858:161-168; Williams et al., 1988, Proc. Natl. Acad. Sci. USA, 85:242-246; Liposomes (Ostro (ed.), 1983, Chapter 1); Hope et al., 1986, Chem. Phys. Lip., 40:89; Gregoriadis, Liposome Technology (1984) and Lasic, Liposomes: from Physics to Applications (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

In certain embodiments, it is desirable to target liposomes using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described (see, e.g., U.S. Pat. Nos. 4,957,773 and 4,603,044).

Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and

carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically
5 recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes lipid components, e.g., phosphatidylethanolamine, which can be activated for attachment of targeting agents, or
10 derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (see Renneisen et al., 1990, J. Biol. Chem., 265:16337-42 and Leonetti et al., 1990, Proc. Natl. Acad. Sci. USA, 87:2448-51).

15 Dosages

For therapeutic applications, the dose of the engineered TALE repeat array protein to be administered to a patient is calculated in a similar way as has been described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent
20 application 2002/0160940. In the context of the present disclosure, the dose should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, e.g., in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy, specificity, and K_D of the particular
25 engineered TALE repeat array protein employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

Pharmaceutical Compositions and Administration

Appropriate pharmaceutical compositions for administration of the engineered TALE repeat array proteins of the present invention can be determined as described for zinc finger proteins, see for example U.S. Pat. No. 6,511,808, U.S. Pat. No. 6,492,117, U.S. Pat. No. 6,453,242, U.S. patent application 2002/0164575, and U.S. patent application 2002/0160940. Engineered TALE repeat array proteins, and expression vectors encoding engineered TALE repeat array proteins, can be administered directly to the patient for modulation of gene expression and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by TALE repeat array protein-mediated gene therapy include pathogenic bacteria, e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria; infectious fungus, e.g., Aspergillus, Candida species; protozoa such as sporozoa (e.g., Plasmodia), rhizopods (e.g., Entamoeba) and flagellates (Trypanosoma, Leishmania, Trichomonas, Giardia, etc.); viral diseases, e.g., hepatitis (A, B, or C), herpes virus (e.g., VZV, HSV-1, HSV-6, HSV-II, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, comovirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, etc.

Administration of therapeutically effective amounts is by any of the routes normally used for introducing TALE repeat array proteins into ultimate contact with the tissue to be treated. The TALE repeat array proteins are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a

particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions that are available (see, e.g., Remington: The Science and Practice of Pharmacy, 21st ed., 2005).

The engineered TALE repeat array proteins, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The disclosed compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

Use of TALE Nucleases

TALE nucleases engineered using the methods described herein can be used to induce mutations in a genomic sequence, e.g., by cleaving at two sites and deleting sequences in between, by cleavage at a single site followed by non-homologous end joining, and/or by cleaving at a site so as to remove or replace one or two or a few nucleotides. In some embodiments, the TALE nuclease is used to induce mutation in an

animal, plant, fungal, or bacterial genome. Targeted cleavage can also be used to create gene knock-outs (e.g., for functional genomics or target validation) and to facilitate targeted insertion of a sequence into a genome (i.e., gene knock-in); e.g., for purposes of cell engineering or protein overexpression. Insertion can be by means of replacements of chromosomal sequences through homologous recombination or by targeted integration, in which a new sequence (i.e., a sequence not present in the region of interest), flanked by sequences homologous to the region of interest in the chromosome, is used to insert the new sequence at a predetermined target site via homologous recombination. Exogenous DNA can also be inserted into TALE nuclease-induced double stranded breaks without the need for flanking homology sequences (see, Orlando et al., 2010, Nucl. Acids Res., 1-15, doi:10.1093/nar/gkq512).

As demonstrated in Example 3 below, the TALE nucleases produced by the methods described herein were capable of inducing site-specific mutagenesis in mammalian cells. A skilled practitioner will readily appreciate that TALE nucleases produced by the methods described herein would also function to induce efficient site-specific mutagenesis in other cell types and organisms (see, for example, Cade et al., 2012, Nucleic Acids Res., PMID: 22684503 and Moore et al., 2012, PLoS One, PMID: 22655075).

The same methods can also be used to replace a wild-type sequence with a mutant sequence, or to convert one allele to a different allele.

Targeted cleavage of infecting or integrated viral genomes can be used to treat viral infections in a host. Additionally, targeted cleavage of genes encoding receptors for viruses can be used to block expression of such receptors, thereby preventing viral infection and/or viral spread in a host organism. Targeted mutagenesis of genes encoding viral receptors (e.g., the CCR5 and CXCR4 receptors for HIV) can be used to render the receptors unable to bind to virus, thereby preventing new infection and blocking the spread of existing infections. Non-limiting examples of viruses or viral receptors that can be targeted include herpes simplex virus (HSV), such as HSV-1 and HSV-2, varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV), HHV6 and HHV7. The hepatitis family of viruses includes hepatitis A virus (HAV), hepatitis B virus

(HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV). Other viruses or their receptors can be targeted, including, but not limited to, Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; Coronaviridae; Reoviridae; 5 Bimaviridae; Rhabdoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviridae; lentiviruses (e.g., HTLV-I; HTLV-II; HIV-1 (also known as HTLV-III, LAV, ARV, hTLR, etc.) HIV-II); simian immunodeficiency virus (SIV), human papillomavirus (HPV), 10 influenza virus and the tick-borne encephalitis viruses. See, e.g., Virology, 3rd Edition (W. K. Joklik, ed. 1988); Fundamental Virology, 4th Edition (Knipe and Howley, eds. 2001), for a description of these and other viruses. Receptors for HIV, for example, include CCR-5 and CXCR-4.

In similar fashion, the genome of an infecting bacterium can be mutagenized by 15 targeted DNA cleavage followed by non-homologous end joining, to block or ameliorate bacterial infections.

The disclosed methods for targeted recombination can be used to replace any genomic sequence with a homologous, non-identical sequence. For example, a mutant genomic sequence can be replaced by its wild-type counterpart, thereby providing 20 methods for treatment of e.g., genetic disease, inherited disorders, cancer, and autoimmune disease. In like fashion, one allele of a gene can be replaced by a different allele using the methods of targeted recombination disclosed herein.

Exemplary genetic diseases include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM 25 No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangiectasia, barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercun's disease, ectodermal dysplasia, Fanconi 30 anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis,

Gaucher's disease, generalized gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefelter's syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920),
5 leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipidus, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID),
10 Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, X-linked lymphoproliferative
15 syndrome (XLP, OMIM No. 308240).

Additional exemplary diseases that can be treated by targeted DNA cleavage and/or homologous recombination include acquired immunodeficiencies, lysosomal storage diseases (e.g., Gaucher's disease, GM1, Fabry disease and Tay-Sachs disease), mucopolysaccharidosis (e.g., Hunter's disease, Hurler's disease), hemoglobinopathies
20 (e.g., sickle cell diseases, HbC, alpha-thalassemia, beta-thalassemia) and hemophilias.

In certain cases, alteration of a genomic sequence in a pluripotent cell (e.g., a hematopoietic stem cell) is desired. Methods for mobilization, enrichment and culture of hematopoietic stem cells are known in the art. See for example, U.S. Pat. Nos. 5,061,620; 5,681,559; 6,335,195; 6,645,489 and 6,667,064. Treated stem cells can be
25 returned to a patient for treatment of various diseases including, but not limited to, SCID and sickle-cell anemia.

In many of these cases, a region of interest comprises a mutation, and the donor polynucleotide comprises the corresponding wild-type sequence. Similarly, a wild-type genomic sequence can be replaced by a mutant sequence, if such is desirable. For
30 example, overexpression of an oncogene can be reversed either by mutating the gene or

by replacing its control sequences with sequences that support a lower, non-pathologic level of expression. As another example, the wild-type allele of the ApoAI gene can be replaced by the ApoAI Milano allele, to treat atherosclerosis. Indeed, any pathology dependent upon a particular genomic sequence, in any fashion, can be corrected or
5 alleviated using the methods and compositions disclosed herein.

Targeted cleavage and targeted recombination can also be used to alter non-coding sequences (e.g., sequences encoding microRNAs and long non-coding RNAs, and regulatory sequences such as promoters, enhancers, initiators, terminators, splice sites) to alter the levels of expression of a gene product. Such methods can be used, for example,
10 for therapeutic purposes, functional genomics and/or target validation studies.

The compositions and methods described herein also allow for novel approaches and systems to address immune reactions of a host to allogeneic grafts. In particular, a major problem faced when allogeneic stem cells (or any type of allogeneic cell) are grafted into a host recipient is the high risk of rejection by the host's immune system,
15 primarily mediated through recognition of the Major Histocompatibility Complex (MHC) on the surface of the engrafted cells. The MHC comprises the HLA class I protein(s) that function as heterodimers that are comprised of a common beta subunit and variable alpha subunits. It has been demonstrated that tissue grafts derived from stem cells that are devoid of HLA escape the host's immune response. See, e.g., Coffman et al., 1993, J.
20 Immunol., 151:425-35; Markmann et al., 1992, Transplantation, 54:1085-89; Koller et al., 1990, Science, 248:1227-30. Using the compositions and methods described herein, genes encoding HLA proteins involved in graft rejection can be cleaved, mutagenized or altered by recombination, in either their coding or regulatory sequences, so that their expression is blocked or they express a non-functional product. For example, by
25 inactivating the gene encoding the common beta subunit gene (beta2 microglobulin) using TALE nuclease fusion proteins as described herein, HLA class I can be removed from the cells to rapidly and reliably generate HLA class I null stem cells from any donor, thereby reducing the need for closely matched donor/recipient MHC haplotypes during stem cell grafting.

Inactivation of any gene (e.g., the beta2 microglobulin gene) can be achieved, for example, by a single cleavage event, by cleavage followed by non-homologous end joining, by cleavage at two sites followed by joining so as to delete the sequence between the two cleavage sites, by targeted recombination of a missense or nonsense codon into the coding region, or by targeted recombination of an irrelevant sequence (i.e., a "stuffer" sequence) into the gene or its regulatory region, so as to disrupt the gene or regulatory region.

Targeted modification of chromatin structure, as disclosed in WO 01/83793, can be used to facilitate the binding of fusion proteins to cellular chromatin.

In additional embodiments, one or more fusions between a TALE binding domain and a recombinase (or functional fragment thereof) can be used, in addition to or instead of the TALE-cleavage domain fusions disclosed herein, to facilitate targeted recombination. See, for example, co-owned U.S. Pat. No. 6,534,261 and Akopian et al. (2003) Proc. Natl. Acad. Sci. USA 100:8688-8691.

In additional embodiments, the disclosed methods and compositions are used to provide fusions of TALE repeat DNA-binding domains with transcriptional activation or repression domains that require dimerization (either homodimerization or heterodimerization) for their activity. In these cases, a fusion polypeptide comprises a TALE repeat DNA-binding domain and a functional domain monomer (e.g., a monomer from a dimeric transcriptional activation or repression domain). Binding of two such fusion polypeptides to properly situated target sites allows dimerization so as to reconstitute a functional transcription activation or repression domain.

Regulation of Gene Expression in Plants

Engineered TALE repeat array proteins can be used to engineer plants for traits such as increased disease resistance, modification of structural and storage polysaccharides, flavors, proteins, and fatty acids, fruit ripening, yield, color, nutritional characteristics, improved storage capability, and the like. In particular, the engineering of crop species for enhanced oil production, e.g., the modification of the fatty acids produced in oilseeds, is of interest.

Seed oils are composed primarily of triacylglycerols (TAGs), which are glycerol esters of fatty acids. Commercial production of these vegetable oils is accounted for primarily by six major oil crops (soybean, oil palm, rapeseed, sunflower, cotton seed, and peanut). Vegetable oils are used predominantly (90%) for human consumption as margarine, shortening, salad oils, and frying oil. The remaining 10% is used for non-food applications such as lubricants, oleochemicals, biofuels, detergents, and other industrial applications.

The desired characteristics of the oil used in each of these applications varies widely, particularly in terms of the chain length and number of double bonds present in the fatty acids making up the TAGs. These properties are manipulated by the plant in order to control membrane fluidity and temperature sensitivity. The same properties can be controlled using TALE repeat array proteins to produce oils with improved characteristics for food and industrial uses.

The primary fatty acids in the TAGs of oilseed crops are 16 to 18 carbons in length and contain 0 to 3 double bonds. Palmitic acid (16:0 [16 carbons: 0 double bonds]), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3) predominate. The number of double bonds, or degree of saturation, determines the melting temperature, reactivity, cooking performance, and health attributes of the resulting oil.

The enzyme responsible for the conversion of oleic acid (18:1) into linoleic acid (18:2) (which is then the precursor for 18:3 formation) is delta-12-oleate desaturase, also referred to as omega-6 desaturase. A block at this step in the fatty acid desaturation pathway should result in the accumulation of oleic acid at the expense of polyunsaturates.

In one embodiment engineered TALE repeat array proteins are used to regulate expression of the FAD2-1 gene in soybeans. Two genes encoding microsomal delta-6 desaturases have been cloned recently from soybean, and are referred to as FAD2-1 and FAD2-2 (Heppard et al., 1996, Plant Physiol. 110:311-319). FAD2-1 (delta-12 desaturase) appears to control the bulk of oleic acid desaturation in the soybean seed. Engineered TALE repeat array proteins can thus be used to modulate gene expression of FAD2-1 in plants. Specifically, engineered TALE repeat array proteins can be used to inhibit expression of the FAD2-1 gene in soybean in order to increase the accumulation

of oleic acid (18:1) in the oil seed. Moreover, engineered TALE proteins can be used to modulate expression of any other plant gene, such as delta-9 desaturase, delta-12 desaturases from other plants, delta-15 desaturase, acetyl-CoA carboxylase, acyl-ACP-thioesterase, ADP-glucose pyrophosphorylase, starch synthase, cellulose synthase, sucrose synthase, senescence-associated genes, heavy metal chelators, fatty acid hydroperoxide lyase, polygalacturonase, EPSP synthase, plant viral genes, plant fungal pathogen genes, and plant bacterial pathogen genes.

Recombinant DNA vectors suitable for transformation of plant cells are also used to deliver protein (e.g., engineered TALE repeat array protein)-encoding nucleic acids to plant cells. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature (see, e.g., Weising et al., 1988, Ann. Rev. Genet., 22:421-477). A DNA sequence coding for the desired TALE repeat array protein is combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the TALE protein in the intended tissues of the transformed plant.

For example, a plant promoter fragment can be employed which will direct expression of the engineered TALE repeat array protein in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35 S transcription initiation region, the 1'- or 2'-promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill.

Alternatively, the plant promoter can direct expression of the engineered TALE repeat array protein in a specific tissue or can be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. Examples of environmental conditions that can affect transcription by inducible promoters include anaerobic conditions or the presence of light.

Examples of promoters under developmental control include promoters that initiate transcription only in certain tissues, such as fruit, seeds, or flowers. For example, the use of a polygalacturonase promoter can direct expression of the TALE repeat array protein in the fruit, a CHS-A (chalcone synthase A from petunia) promoter can direct
5 expression of the TALE repeat array protein in the flower of a plant.

The vector comprising the TALE repeat array protein sequences will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker can encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such
10 as resistance to chlorosulfuron or Basta.

Such DNA constructs can be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be
15 introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment. Alternatively, the DNA constructs can be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is
20 infected by the bacteria.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski et al., 1984, EMBO J., 3:2717-22.

Electroporation techniques are described in Fromm et al. 1985, Proc. Natl. Acad. Sci. USA, 82:5824. Biolistic transformation techniques are described in Klein et al., 1987,
25 Nature, 327:70-73.

Agrobacterium tumefaciens-mediated transformation techniques are well described in the scientific literature (see, e.g., Horsch et al., 1984, Science, 233:496-498; and Fraley et al., 1983, Proc. Natl. Acad. Sci. USA, 80:4803).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired TALE repeat array protein-controlled phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the TALE repeat array protein nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176 (1983); and Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73 (1985). Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al., 1987, Ann. Rev. Plant Phys., 38:467-486.

Functional Genomics Assays

Engineered TALE repeat array proteins also have use for assays to determine the phenotypic consequences and function of gene expression. Recent advances in analytical techniques, coupled with focused mass sequencing efforts have created the opportunity to identify and characterize many more molecular targets than were previously available. This new information about genes and their functions will improve basic biological understanding and present many new targets for therapeutic intervention. In some cases analytical tools have not kept pace with the generation of new data. An example is provided by recent advances in the measurement of global differential gene expression. These methods, typified by gene expression microarrays, differential cDNA cloning frequencies, subtractive hybridization and differential display methods, can very rapidly identify genes that are up or down-regulated in different tissues or in response to specific stimuli. Increasingly, such methods are being used to explore biological processes such as, transformation, tumor progression, the inflammatory response, neurological disorders etc. Many differentially expressed genes correlate with a given physiological phenomenon, but demonstrating a causative relationship between an individual differentially expressed gene and the phenomenon is labor intensive. Until now, simple

methods for assigning function to differentially expressed genes have not kept pace with the ability to monitor differential gene expression.

The engineered TALE repeat array proteins described herein can be used to rapidly analyze the function of a differentially expressed gene. Engineered TALE proteins can be readily used to up or down-regulate or knockout any endogenous target gene, or to knock in an endogenous or endogenous gene. Very little sequence information is required to create a gene-specific DNA binding domain. This makes the engineered TALE repeat array technology ideal for analysis of long lists of poorly characterized differentially expressed genes. One can simply build a TALE repeat array protein-based DNA binding domain for each candidate gene, create chimeric up and down-regulating artificial transcription factors and test the consequence of up or down-regulation on the phenotype under study (e.g., transformation or response to a cytokine) by switching the candidate genes on or off one at a time in a model system.

Additionally, greater experimental control can be imparted by engineered TALE repeat array proteins than can be achieved by more conventional methods. This is because the production and/or function of engineered TALE repeat array proteins can be placed under small molecule control. Examples of this approach are provided by the Tet-On system, the ecdysone-regulated system and a system incorporating a chimeric factor including a mutant progesterone receptor. These systems are all capable of indirectly imparting small molecule control on any endogenous gene of interest or any transgene by placing the function and/or expression of a engineered TALE repeat array protein under small molecule control.

Transgenic Animals

A further application of engineered TALE repeat array proteins is manipulating gene expression in animal models. As with cell lines, the introduction of a heterologous gene into or knockout of an endogenous in a transgenic animal, such as a transgenic mouse or zebrafish, is a fairly straightforward process. Thus, transgenic or transient expression of an engineered TALE repeat array protein in an animal can be readily performed.

By transgenically or transiently expressing a suitable engineered TALE repeat array protein fused to an activation domain, a target gene of interest can be over-expressed. Similarly, by transgenically or transiently expressing a suitable engineered TALE repeat array protein fused to a repressor or silencer domain, the expression of a target gene of interest can be down-regulated, or even switched off to create "functional knockout". Knock-in or knockout mutations by insertion or deletion of a target gene of interest can be prepared using TALE nucleases.

Two common issues often prevent the successful application of the standard transgenic and knockout technology; embryonic lethality and developmental compensation. Embryonic lethality results when the gene plays an essential role in development. Developmental compensation is the substitution of a related gene product for the gene product being knocked out, and often results in a lack of a phenotype in a knockout mouse when the ablation of that gene's function would otherwise cause a physiological change.

Expression of transgenic engineered TALE repeat array proteins can be temporally controlled, for example using small molecule regulated systems as described in the previous section. Thus, by switching on expression of an engineered TALE repeat array protein at a desired stage in development, a gene can be over-expressed or "functionally knocked-out" in the adult (or at a late stage in development), thus avoiding the problems of embryonic lethality and developmental compensation.

EXAMPLES

Example 1. Assembly of TALE Repeat Arrays using streptavidin coated magnetic beads

An archive of DNA plasmids (~850 different plasmids) encoding one, two, three, or four TALE repeat domains was created for assembly of nucleic acids encoding multiple TALE arrays of any desired length. The plasmids were created by cloning synthetic arrays of one, two, three or four TALE repeat domains into the pUC57- Δ Bsal backbone (FIG. 3). The TALE repeats were of the arrangement α , $\beta\gamma\delta\epsilon$, $\beta\gamma\delta$, $\beta\gamma'$, $\beta\gamma$, $\delta\epsilon'$, and β , and included hypervariable triplet residues at each position to bind to the nucleotides as shown in Table 1. Polypeptide and nucleotide sequences of the TALE

repeat types are shown in FIGs. 4A and 4B, respectively. The polypeptide and polynucleotide sequences were varied slightly among the four types to reduce the possibility of recombination-mediated mutations due to long sequences of exact repeats.

5

Table 1. Nucleotide binding code of TALE triplets

Triplet	Bound Nucleotide
SNI	A
SHD	C
NNN	G
SNK	G
SNG	T

A 16-mer TALE repeat array targeted to the eGFP gene was created by in vitro assembly of 16 TALE repeats designed to bind the target sequence GCAGTGCTTCAGCCGC (SEQ ID NO: 41). In the first step, a plasmid carrying an α -type TALE repeat with an NNN triplet (G) was amplified by PCR using a biotinylated forward primer Biotin-TCTAGAGAAGACAAGAACCTGACC (SEQ ID NO: 42) and a reverse primer GGATCCGGTCTCTTAAGGCCGTGG (SEQ ID NO: 43). The amplified fragment (50 μ l) was purified using a QIA Quick PCR purification kit (QIAGEN), eluted in 40 μ l 0.1X elution buffer (as provided in the QIA Quick PCR purification kit), and digested with BsaI HF (New England Biolabs (NEB)) in NEB Buffer 4 for 15 minutes at 50 °C (40 μ l elution, 5 μ l NEBuffer 4, 5 μ l BsaI HF). The digested fragment was purified using a QIA Quick PCR purification kit and eluted in 0.1X elution buffer (50 μ l).

A plasmid containing a four TALE repeat domain sub-array unit ($\beta\gamma\delta\epsilon$) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG (designed to bind the sequence 5'-CAGT-3') was digested with BbsI (NEB) in NEBuffer 2 for 2 hours at 37 °C in 100 μ l (50 μ l plasmid [$\sim$ 200 ng/ μ l], 10 μ l NEBuffer 2, 10 μ l BbsI, 30 μ l water). To the 100 μ l digest was added 25 μ l NEBuffer 4, 2.5 μ l 100X BSA (NEB), 107.5 μ l water, and 5 μ l XbaI (NEB), and the digest was incubated for 5

minutes at 37 °C. To the mixture, 5 µl of BamHI HF was then added for a 5 minute digest at 37 °C, and then 5 µl Sall HF (NEB) was added for an additional 5 minute digest at 37 °C. The resulting fragment was purified using a QIA Quick PCR purification kit (QIAGEN) and eluted in 180 µl 0.1X elution buffer.

5 For the initial ligation, 2 µl of the alpha unit digest was mixed with 2.5 µl of T4 DNA ligase (400 U/µl; NEB) and 27 µl Quick Ligase Buffer (QLB) (NEB). To this 31.5 µl mixture was added 22.5 µl of the first digested subarray, and the mixture was ligated for 15 minutes at room temperature. Magnetic beads were prepared by washing 5 µl of Dynabeads MyOne Streptavidin C1 (Invitrogen) three times with 50 µl 1x B&W Buffer (5.0 mM Tris-HCl [pH 7.5], 0.5 mM EDTA, 1.0 M NaCl, 0.005% Tween 20) and
10 resuspending in 54 µl B&W Buffer. The ligated mixture was added to the washed beads and incubated for 15 minutes at room temperature (with mixing every five minutes). The mixture was then placed on a SPRIplate 96-well Ring magnet for 3 minutes. The supernatant was then aspirated, and 100 µl 1x B&W Buffer was added to wash, with
15 mixing by moving the beads 31 times from side to side within the tube using a DynaMag-96 Side magnet (Invitrogen). The B&W Buffer was then aspirated, and 100 µl 1x BSA was added, with mixing, then aspirated. The ligated, bead-bound nucleic acids ($\alpha\beta\gamma\delta\epsilon$) were resuspended in 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water).

 The digest was incubated at 50 °C for 10 minutes, and 50 µl 1X B&W buffer was
20 added. The digest was placed on a magnet for 3 minutes, and the supernatant was aspirated. The beads were washed with 100 µl 1X B&W Buffer and 100 µl 1X BSA as above. To the washed beads were added a digested plasmid containing a four TALE repeat domain sub-array unit ($\beta\gamma\delta\epsilon$) coding for repeats that each harbor one of the following variable amino acids NNN, SHD, SNG, and SNG (designed to bind the DNA
25 sequence 5'-GCTT-3') (22.5 µl) and 27.5 µl ligase mix (25 µl Quick Ligase Buffer, 2 µl DNA ligase). The beads were resuspended by pipetting up and down, and the mixture was incubated for 15 minutes at room temperature with mixing every five minutes. To the ligation was added 50 µl 1X B&W Buffer, and the mixture was placed on the magnet for 3 minutes. The supernatant was aspirated, and the beads were washed with 100 µl 1X
30 B&W Buffer and 100 µl 1X BSA as above. The ligated, bead-bound nucleic acids

($\alpha\beta\gamma\delta\epsilon\beta\gamma\delta\epsilon$) were resuspended in 50 μ l BsaI HF mix (5 μ l NEBuffer 4, 2 μ l BsaI HF, 43 μ l water). Two more TALE repeat sub-array units were ligated sequentially as above, the first a four TALE repeat sub-array unit ($\beta\gamma\delta\epsilon$) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SHD (designed to bind the DNA sequence 5'-CAGC-3') and the second a three TALE repeat sub-array unit ($\beta\gamma\delta$) coding for repeats that each harbor one of the following variable amino acids SHD, NNN, and SHD (designed to bind the DNA sequence 5'-CGC-3'). The final TALE repeat array contained subunits of the format $\alpha\beta\gamma\delta\epsilon\beta\gamma\delta\epsilon\beta\gamma\delta$ with individual TALE repeats designed to bind the target DNA sequence 5'-GCAGTGCTTCAGCCGC-3' (SEQ ID NO: 44).

Following the final ligation step, the construct was digested with BsaI HF for eventual cloning into an expression vector and the beads were washed with 1X B&W Buffer and 1X BSA. The washed beads were resuspended in 50 μ l BbsI mix (5 μ l NEBuffer 2, 5 μ l BbsI, 40 μ l water) and incubated at 37 °C for 2 hours with agitation at 1500 rpm to cleave the biotinylated 5' end and release the assembled TALE repeat array from the magnetic beads. The digested mixture was purified by MinElute column purified (QIAGEN) and ligated into a BsmBI-digested TALE expression vector. The ligated mixture was transformed into chemically competent XL1 Blue cells and plated on LB/Carb¹⁰⁰ plates overnight.

The expression vectors each harbor the following elements: a T7 promoter, a nuclear localization signal, a FLAG tag, amino acids 153 to 288 from the TALE13 protein (numbering as defined by Miller et al., 2011, Nat. Biotechnol., 29:143-148), two adjacent BsmBI restriction sites into which a DNA fragment encoding a TALE repeat array can be cloned, a 0.5 TALE repeat, amino acids 715 to 777 from the C-terminal end of the TALE13 protein (numbering as defined by Miller et al., 2011, Nat. Biotechnol., 29:143-148), and the wild-type FokI cleavage domain.

The plasmids differ in the identity of the C-terminal 0.5 TALE repeat. Plasmid pJDS70 encodes a 0.5 TALE repeat with a SNI RVD (for recognition of an A nucleotide), plasmid pJDS71 encodes a 0.5 TALE repeat with a SHD RVD (for recognition of a C nucleotide), plasmid pJDS74 encodes a 0.5 TALE repeat with a NNN RVD (for

recognition of a G nucleotide), plasmid pJDS76 encodes a 0.5 TALE repeat with a SNK RVD (for recognition of a G nucleotide), and plasmid pJDS78 encodes a 0.5 TALE repeat with a NG RVD (for recognition of a T nucleotide). All plasmids share the common sequence shown in FIGs. 5A-5B and differ at just nine nucleotide positions marked as
 5 XXXXXXXXXX (underlined and bold). The sequence of these 9 bps and plasmid names are also shown below in Table 2.

Table 2. DNA sequences of expression vectors

Plasmid name	Sequence of variable 9 bps	SEQ ID NO:	RVD of C-terminal 0.5 TALE repeat
pJDS70	TCTAACATC	45	SNI (for binding to an A nucleotide)
pJDS71	TCCCACGAC	46	SHD (for binding to a C nucleotide)
pJDS74	AATAATAAC	47	NNN (for binding to a G nucleotide)
pJDS76	TCCAATAAA	48	SNK (for binding to a G nucleotide)
pJDS78	TCTAATGGG	49	SNG (for binding to a T nucleotide)

10 This example demonstrates the construction of TALE repeat arrays on an immobilized substrate using preassembled TALE repeat sub-array units. The above method, up to the cloning step, can be performed in one day.

Example 2. Assembly of TALE Repeat Arrays using a streptavidin coated plate

15 TALE repeats are assembled using the archive of DNA plasmids (~850 different plasmids) as described in Example 1. A 16-mer TALE repeat array was created by in vitro assembly of 16 TALE repeats designed to bind a target sequence. In the first step, a plasmid carrying an α -type TALE repeat with an NNN triplet (G) was amplified by PCR using a biotinylated forward primer Biotin-TCTAGAGAAGACAAGAACCTGACC
 20 (SEQ ID NO: 42) and a reverse primer GGATCCGGTCTCTTAAGGCCGTGG (SEQ ID NO: 43). The amplified fragment (50 μ l) was purified using a QIA Quick PCR purification kit (QIAGEN), eluted in 40 μ l 0.1X elution buffer (as provided in the QIA

Quick PCR purification kit), and digested with BsaI HF (New England Biolabs (NEB)) in NEB Buffer 4 for 15 minutes at 50 °C (40 µl elution, 5 µl NEBuffer 4, 5 µl BsaI HF).

The digested fragment was purified using a QIA Quick PCR purification kit and eluted in 0.1X elution buffer (50 µl).

5 A plasmid containing a four TALE repeat domain sub-array unit ($\beta\gamma\delta\epsilon$) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG (designed to bind the sequence 5'-CAGT-3') was digested with BbsI (NEB) in NEBuffer 2 for 2 hours at 37 °C in 100 µl (50 µl plasmid [~ 200 ng/ μ l], 10 µl NEBuffer 2, 10 µl BbsI, 30 µl water). To the 100 µl digest was added 25 µl NEBuffer 4, 2.5 µl 100X BSA (NEB), 107.5 µl water, and 5 µl XbaI (NEB), and the digest was incubated for 5
10 minutes at 37 °C. To the mixture, 5 µl of BamHI HF was then added for a 5 minute digest at 37 °C, and then 5 µl SalI HF (NEB) was added for an additional 5 minute digest at 37 °C. The resulting fragment was purified using a QIA Quick PCR purification kit (QIAGEN) and eluted in 180 µl 0.1X elution buffer.

15 For the initial ligation, 2 µl of the alpha unit digest was mixed with 2.5 µl of T4 DNA ligase (400 U/ μ l; NEB) and 27 µl Quick Ligase Buffer (QLB) (NEB). To this 31.5 µl mixture was added 22.5 µl of the first digested subarray, and the mixture was ligated for 15 minutes at room temperature. The ligation mixture was then mixed with 2X B&@ buffer (Invitrogen) and added to a well in a 96-well plate coated with
20 streptavidin (Thermo Scientific) and incubated at room temperature for 15min. The supernatant was aspirated. Each well in the 96 well plate was washed with 200ul of 1x Bovine Serum Albumin (BSA) by pipetting up and down 10 times before discarding the 1x BSA. This was repeated for a total of two washes with 1x BSA. Then 50 µl BsaI HF mix (5 µl NEBuffer 4, 2 µl BsaI HF, 43 µl water) was added to the ligated, nucleic acids
25 ($\alpha\beta\gamma\delta\epsilon$) bound to the streptavidin-coated well.

 The digest was incubated at 50 °C for 10 minutes and then the supernatant was aspirated. The wells were then washed with 200 µl 1X B&W Buffer and 200 µl 1X BSA twice by pipetting up and down ten times before removal of each supernatant. 22.5 µl of digested plasmid encoding a four TALE repeat domain sub-array unit ($\beta\gamma\delta\epsilon$) coding for
30 repeats that each harbor one of the following variable amino acids NNN, SHD, SNG, and

SNI and 27.5 μ l ligase mix (25 μ l Quick Ligase Buffer, 2 μ l DNA ligase) were added to the well. The supernatant was mixed by pipetting up and down, and the mixture was incubated for 15 minutes at room temperature. The supernatant was removed and the well was washed with 1X B&W and 1x BSA as above. Then 50 μ l BsaI HF mix (5 μ l
 5 NEBuffer 4, 2 μ l BsaI HF, 43 μ l water) was added to the ligated nucleic acids ($\alpha\beta\gamma\delta\epsilon\beta\gamma\delta\epsilon$) bound to the well. Two more TALE repeat sub-array units were ligated sequentially as above, the first a four TALE repeat sub-array unit ($\beta\gamma\delta\epsilon$) coding for repeats that each harbor one of the following variable amino acids SHD, SNI, NNN, and SNG and the second a three TALE repeat sub-array unit ($\beta\gamma\delta$) coding for repeats that
 10 each harbor one of the following variable amino acids SHD, SNI, NNN, and SHD. The final TALE repeat array contained subunits of the format $\alpha\beta\gamma\delta\epsilon\beta\gamma\delta\epsilon\beta\gamma\delta$ with individual TALE repeats designed to bind a target DNA sequence.

Following the final ligation step, the fragments in the well were digested with BsaI HF for eventual cloning into an expression vector. The well was then washed with
 15 1X B&W Buffer and twice with 1X BSA. Then 50 μ l BbsI mix (5 μ l NEBuffer 2, 5 μ l BbsI, 40 μ l water) was added to the well and incubated at 37 °C for 2 hours to cleave the biotinylated 5' end and release the assembled TALE repeat array from the well. The digested mixture was purified, ligated, and transformed as described in Example 1.

20 Example 3. Site-Specific Mutagenesis Using TALE Nucleases

To demonstrate the effectiveness of TALE repeat domains created by the methods described herein, TALE repeat arrays were constructed and cloned into TALE nuclease expression vectors (as described in Example 1) to produce plasmids encoding TALE
 nuclease monomers targeted to the eGFP coding sequences shown in FIG. 6 and Table 3.
 25 Nucleic acid and polypeptide sequences of the TALE nuclease monomers are shown in FIGs. 11A-18B.

Table 3. TALE nuclease monomer target sequences

TALE Fragment	Target Sequence	Length of target sequence	SEQ ID NO:	Site	Position (half-site)	Plasmid name
DR-TALE-0003	TGCAGTGCTTCAGCCGC	17	50	eGFP223	left	SQT70
DR-TALE-0006	TGCAGTGCTTCAGCCGCT	18	51	eGFP223	left	SQT114
DR-TALE-0005	TTGAAGAAGTCGTGCTGC	18	52	eGFP223	right	SQT72
DR-TALE-0010	TGAAGAAGTCGTGCTGCT	18	53	eGFP223	right	SQT56
DR-TALE-0023	TCGAGCTGAAGGGCATC	17	54	eGFP382	left	SQT84
DR-TALE-0025	TCGAGCTGAAGGGCATCG	18	55	eGFP382	left	SQT120
DR-TALE-0020	TTGTGCCCCAGGATGTTG	18	56	eGFP382	right	SQT135
DR-TALE-0022	TGTGCCCCAGGATGTTGC	18	57	eGFP382	right	SQT118

4E5 U2OS-eGFP cells were nucleofected with 400 ng plasmid DNA in solution SE with program DN-100 using Nucleofector™ non-viral transfection (Lonza, Walkersville, MD). The cells were analyzed by flow cytometry at days 2 and 5 (FIG. 7). Non-homologous end joining (NHEJ)-mediated mutagenic repair of TALE nuclease-induced double-stranded breaks led to disruption of eGFP expression (eGFP-negative cells). All eight TALE nuclease pairs tested induced a high percentage of eGFP-negative (eGFP-) cells (y-axis). The percentage of eGFP- cells declined only modestly between day 2 and 5 suggesting that the alterations were stably induced.

A subset of mutated eGFP genes were amplified from cells and sequenced. The resulting mutations are shown in FIG. 8. Sequences targeted by the TALE nucleases encoded by expression plasmids SQT70/SQT56 in human USOS-eGFP cells are

underlined in the wild-type (WT) sequence shown at the top of FIG. 8. Insertion and deletion mutations induced by the TALE nuclease pair are shown below with deleted bases indicated by dashes and inserted bases indicated by double underlining. The net number of bases inserted or deleted is shown to the right. All mutations were isolated once unless otherwise indicated in brackets. The overall frequency of mutagenesis (46%) is also indicated.

Example 4. Automated Assembly of TALE Repeat Arrays

The assembly method described in Example 1 has been automated so as to be performed using a Sciclone™ G3 liquid handling workstation (Caliper Life Sciences, Hopkinton, MA) in 96-well plates. All of the steps were automated except digestion of the nucleic acids prior to ligation and linking to the beads and the steps following release of the assembled TALE repeat array from the magnetic beads. The automated steps were performed essentially as when done manually with minor variations in the number of resuspension and mixing motions. The results of assembly of two 17-mers are shown in FIG. 9. A major product of the expected size can be seen, corresponding to the 17-mer. Additional minor 13-mer, 9-mer, and 5-mer products can also be seen, likely produced by carry forward of incompletely ligated products. A similar result can be seen in FIG. 10, which shows the results of assembly of 16-mers from an N-terminal 1-mer sub-array (1), three 4-mer subarrays (4_A, 4_B, 4_C), and a C-terminal 3-mer subarray (3_D).

This example demonstrates that the methods described herein can be automated for rapid and reproducible synthesis of nucleic acids encoding TALE repeat arrays.

Example 5. Assembly Methods

TALE repeat arrays were created using an architecture in which four distinct TALE repeat backbones that differ slightly in their amino acid and DNA sequences occur in a repeated pattern. The first, amino-terminal TALE repeat in an array was designated as the α unit. This was followed by β , γ , and δ units and then an ϵ unit that is essentially identical to the α unit except for the different positioning of a Type IIS restriction site on the 5' end (required to enable creation of a unique overhang on the α unit needed for

cloning). The ϵ unit was then followed again by repeats of β , γ , δ , and ϵ units. Due to constraints related to creation of a 3' end required for cloning, slightly modified DNA sequences were required for TALE repeat arrays that end with a carboxy-terminal γ or ϵ unit. We designated these variant units as γ^* and ϵ^* .

5 For each type of TALE repeat unit (i.e.— α , β , γ , δ , ϵ , γ^* , and ϵ^*), we commercially synthesized (Genscript) a series of four plasmids, each harboring one of the five repeat variable di-residues (RVDs) that specifies one of the four DNA bases (NI = A; HD = C; NN = G; NG = T, NK = G). Full DNA sequences of these plasmids are provided in Table 4 and FIG. 3. For all 35 of these plasmids, the sequence encoding the
10 TALE repeat domain is flanked on the 5' end by unique XbaI and BbsI restriction sites and on the 3' end by unique BsaI and BamHI restriction sites. Additionally, the overhangs generated by digestion of any plasmids encoding units designed to be adjacent to one another (e.g.— β and γ , or δ and ϵ) with BsaI and BbsI are complementary. Using these 35 different plasmids and serial ligation via the BsaI and BbsI restriction sites, we
15 assembled an archive of all possible combinations of $\beta\gamma$, $\beta\gamma\delta\epsilon$, $\beta\gamma\delta$, $\beta\gamma^*$, and $\delta\epsilon^*$ repeats. In total, this archive consisted of 825 different plasmids encoding 5 α 's, 5 β 's, 25 $\beta\gamma$ combinations, 625 $\beta\gamma\delta\epsilon$ combinations, 125 $\beta\gamma\delta$ combinations, 25 $\beta\gamma^*$ combinations, and 25 $\delta\epsilon^*$ combinations (Table 5). These 825 plasmids plus ten of the original 35 plasmids encoding single TALE repeats (five α and five β plasmids) are required to practice the
20 methods. With this archive of 835 plasmids listed in Table 5, the methods can be used to construct TALE repeat arrays of any desired length and composition.

Table 4. DNA sequences encoding individual TALE repeats

TAL ID#	Unit Architecture	RVD	Target Base	DNA Sequence (Cloned between XbaI/BamHI in pUC57- Δ BsaI)	SEQ ID NO:
6	α	NI	A	TCTAGAGAAGACAAGAACCCTGACC CCAGACCAGGTAGTCGCAATCGCG TCGAACATTGGGGGAAAGCAAGCC CTGGAAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC	58

7	α	HD	C	TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCACATGACGGGGGAAAGCAAGCC CTGGAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC	59
8	α	NK	G	TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCGAACAAAGGGGGAAAGCAAGCC CTGGAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC	60
9	α	NN	G	TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG AACAATAATGGGGGAAAGCAAGCC CTGGAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC	61
10	α	NG	T	TCTAGAGAAGACAAGAACCTGACC CCAGACCAGGTAGTCGCAATCGCG TCAAACGGAGGGGGAAAGCAAGCC CTGGAACCGTGCAAAGGTTGTTG CCGGTCCTTTGTCAAGACCACGGC CTTAAGAGACCGGATCC	62
11	β	NI	A	TCTAGAGAAGACAACCTTACACCGG AGCAAGTCGTGGCCATTGCAAGCA ACATCGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC	63
12	β	HD	C	TCTAGAGAAGACAACCTTACACCGG AGCAAGTCGTGGCCATTGCATCCC ACGACGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC	64
13	β	NK	G	TCTAGAGAAGACAACCTTACACCGG AGCAAGTCGTGGCCATTGCATCAA ATAAAGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC	65
14	β	NN	G	TCTAGAGAAGACAACCTTACACCGG AGCAAGTCGTGGCCATTGCAAATA ATAACGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC	66

15	β	NG	T	TCTAGAGAAGACAACCTTACACCGG AGCAAGTCGTGGCCATTGCAAGCA ATGGGGGTGGCAAACAGGCTCTTG AGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGGCTGA AGAGACCGGATCC	67
16	γ	NI	A	TCTAGAGAAGACAACCTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGA ACATTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC	68
17	γ	HD	C	TCTAGAGAAGACAACCTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGC ATGACGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC	69
18	γ	NK	G	TCTAGAGAAGACAACCTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACAAGGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC	70
19	γ	NN	G	TCTAGAGAAGACAACCTGACTCCCG ATCAAGTTGTAGCGATTGCGAATA ACAATGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC	71
20	γ	NG	T	TCTAGAGAAGACAACCTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACGGTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTTTGA AGAGACCGGATCC	72
21	δ	NI	A	TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCTCCA ATATTGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC	73
22	δ	HD	C	TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAGCC ATGATGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC	74

23	δ	NK	G	TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAGCA ATAAGGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC	75
24	δ	NN	G	TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCAACA ACAACGGCGGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC	76
25	δ	NG	T	TCTAGAGAAGACAATTGACGCCTG CACAAGTGGTCGCCATCGCCTCGA ATGGCGGGCGTAAGCAGGCGCTGG AAACAGTACAGCGCCTGCTGCCTG TACTGTGCCAGGATCATGGACTGA AGAGACCGGATCC	77
26	ϵ	NI	A	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGTCGA ACATTGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC	78
27	ϵ	HD	C	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGTCAC ATGACGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC	79
28	ϵ	NK	G	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGTCGA ACAAAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC	80
29	ϵ	NN	G	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGAACA ATAATGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC	81
30	ϵ	NG	T	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGTCAA ACGGAGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTTA AGAGACCGGATCC	82

31	γ'	NI	A	TCTAGAGAAGACAACACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGA ACATTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC	83
32	γ'	HD	C	TCTAGAGAAGACAACACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCGC ATGACGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC	84
33	γ'	NK	G	TCTAGAGAAGACAACACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACAAGGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC	85
34	γ'	NN	G	TCTAGAGAAGACAACACTGACTCCCG ATCAAGTTGTAGCGATTGCGAATA ACAATGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC	86
35	γ'	NG	T	TCTAGAGAAGACAACACTGACTCCCG ATCAAGTTGTAGCGATTGCGTCCA ACGGTGGAGGGAAACAAGCATTGG AGACTGTCCAACGGCTCCTTCCCG TGTTGTGTCAAGCCCACGGTCTGA AGAGACCGGATCC	87
36	ϵ'	NI	A	TCTAGAGAAGACAACACTGACCCAG ACCAGGTAGTCGCAATCGCGTCGA ACATTGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC	88
37	ϵ'	HD	C	TCTAGAGAAGACAACACTGACCCAG ACCAGGTAGTCGCAATCGCGTAC ATGACGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC	89
38	ϵ'	NK	G	TCTAGAGAAGACAACACTGACCCAG ACCAGGTAGTCGCAATCGCGTCGA ACAAAGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC	90

39	ϵ'	NN	G	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGAACA ATAATGGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC	91
40	ϵ'	NG	T	TCTAGAGAAGACAAGTACCCAG ACCAGGTAGTCGCAATCGCGTCAA ACGGAGGGGAAAGCAAGCCCTGG AAACCGTGCAAAGGTTGTTGCCGG TCCTTTGTCAAGACCACGGCCTGA AGAGACCGGATCC	92

Table 5. Archive of 835 plasmids encoding pre-assembled TALE repeat units

Plasmid ID	DNA Target	RVDs	Unit Architecture
TAL006	A	NI	α
TAL007	C	HD	α
TAL008	G	NK	α
TAL009	G	NN	α
TAL010	T	NG	α
TAL011/016/021/026	AAAA	NI/NI/NI/NI	$\beta\gamma\delta\epsilon$
TAL011/016/021/027	AAAC	NI/NI/NI/HD	$\beta\gamma\delta\epsilon$
TAL011/016/021/028	AAAG	NI/NI/NI/NK	$\beta\gamma\delta\epsilon$
TAL011/016/021/029	AAAG	NI/NI/NI/NN	$\beta\gamma\delta\epsilon$
TAL011/016/021/030	AAAT	NI/NI/NI/NG	$\beta\gamma\delta\epsilon$
TAL011/016/022/026	AACA	NI/NI/HD/NI	$\beta\gamma\delta\epsilon$
TAL011/016/022/027	AACC	NI/NI/HD/HD	$\beta\gamma\delta\epsilon$
TAL011/016/022/028	AACG	NI/NI/HD/NK	$\beta\gamma\delta\epsilon$
TAL011/016/022/029	AACG	NI/NI/HD/NN	$\beta\gamma\delta\epsilon$
TAL011/016/022/030	AACT	NI/NI/HD/NG	$\beta\gamma\delta\epsilon$
TAL011/016/023/026	AAGA	NI/NI/NK/NI	$\beta\gamma\delta\epsilon$
TAL011/016/023/027	AAGC	NI/NI/NK/HD	$\beta\gamma\delta\epsilon$
TAL011/016/023/028	AAGG	NI/NI/NK/NK	$\beta\gamma\delta\epsilon$
TAL011/016/023/029	AAGG	NI/NI/NK/NN	$\beta\gamma\delta\epsilon$
TAL011/016/023/030	AAGT	NI/NI/NK/NG	$\beta\gamma\delta\epsilon$
TAL011/016/024/026	AAGA	NI/NI/NN/NI	$\beta\gamma\delta\epsilon$
TAL011/016/024/027	AAGC	NI/NI/NN/HD	$\beta\gamma\delta\epsilon$
TAL011/016/024/028	AAGG	NI/NI/NN/NK	$\beta\gamma\delta\epsilon$
TAL011/016/024/029	AAGG	NI/NI/NN/NN	$\beta\gamma\delta\epsilon$
TAL011/016/024/030	AAGT	NI/NI/NN/NG	$\beta\gamma\delta\epsilon$
TAL011/016/025/026	AATA	NI/NI/NG/NI	$\beta\gamma\delta\epsilon$

TAL011/016/025/027	AATC	NI/NI/NG/HD	βγδε
TAL011/016/025/028	AATG	NI/NI/NG/NK	βγδε
TAL011/016/025/029	AATG	NI/NI/NG/NN	βγδε
TAL011/016/025/030	AATT	NI/NI/NG/NG	βγδε
TAL011/017/021/026	ACAA	NI/HD/NI/NI	βγδε
TAL011/017/021/027	ACAC	NI/HD/NI/HD	βγδε
TAL011/017/021/028	ACAG	NI/HD/NI/NK	βγδε
TAL011/017/021/029	ACAG	NI/HD/NI/NN	βγδε
TAL011/017/021/030	ACAT	NI/HD/NI/NG	βγδε
TAL011/017/022/026	ACCA	NI/HD/HD/NI	βγδε
TAL011/017/022/027	ACCC	NI/HD/HD/HD	βγδε
TAL011/017/022/028	ACCG	NI/HD/HD/NK	βγδε
TAL011/017/022/029	ACCG	NI/HD/HD/NN	βγδε
TAL011/017/022/030	ACCT	NI/HD/HD/NG	βγδε
TAL011/017/023/026	ACGA	NI/HD/NK/NI	βγδε
TAL011/017/023/027	ACGC	NI/HD/NK/HD	βγδε
TAL011/017/023/028	ACGG	NI/HD/NK/NK	βγδε
TAL011/017/023/029	ACGG	NI/HD/NK/NN	βγδε
TAL011/017/023/030	ACGT	NI/HD/NK/NG	βγδε
TAL011/017/024/026	ACGA	NI/HD/NN/NI	βγδε
TAL011/017/024/027	ACGC	NI/HD/NN/HD	βγδε
TAL011/017/024/028	ACGG	NI/HD/NN/NK	βγδε
TAL011/017/024/029	ACGG	NI/HD/NN/NN	βγδε
TAL011/017/024/030	ACGT	NI/HD/NN/NG	βγδε
TAL011/017/025/026	ACTA	NI/HD/NG/NI	βγδε
TAL011/017/025/027	ACTC	NI/HD/NG/HD	βγδε
TAL011/017/025/028	ACTG	NI/HD/NG/NK	βγδε
TAL011/017/025/029	ACTG	NI/HD/NG/NN	βγδε
TAL011/017/025/030	ACTT	NI/HD/NG/NG	βγδε
TAL011/018/021/026	AGAA	NI/NK/NI/NI	βγδε
TAL011/018/021/027	AGAC	NI/NK/NI/HD	βγδε
TAL011/018/021/028	AGAG	NI/NK/NI/NK	βγδε
TAL011/018/021/029	AGAG	NI/NK/NI/NN	βγδε
TAL011/018/021/030	AGAT	NI/NK/NI/NG	βγδε
TAL011/018/022/026	AGCA	NI/NK/HD/NI	βγδε
TAL011/018/022/027	AGCC	NI/NK/HD/HD	βγδε
TAL011/018/022/028	AGCG	NI/NK/HD/NK	βγδε
TAL011/018/022/029	AGCG	NI/NK/HD/NN	βγδε
TAL011/018/022/030	AGCT	NI/NK/HD/NG	βγδε
TAL011/018/023/026	AGGA	NI/NK/NK/NI	βγδε
TAL011/018/023/027	AGGC	NI/NK/NK/HD	βγδε

TAL011/018/023/028	AGGG	NI/NK/NK/NK	βγδε
TAL011/018/023/029	AGGG	NI/NK/NK/NN	βγδε
TAL011/018/023/030	AGGT	NI/NK/NK/NG	βγδε
TAL011/018/024/026	AGGA	NI/NK/NN/NI	βγδε
TAL011/018/024/027	AGGC	NI/NK/NN/HD	βγδε
TAL011/018/024/028	AGGG	NI/NK/NN/NK	βγδε
TAL011/018/024/029	AGGG	NI/NK/NN/NN	βγδε
TAL011/018/024/030	AGGT	NI/NK/NN/NG	βγδε
TAL011/018/025/026	AGTA	NI/NK/NG/NI	βγδε
TAL011/018/025/027	AGTC	NI/NK/NG/HD	βγδε
TAL011/018/025/028	AGTG	NI/NK/NG/NK	βγδε
TAL011/018/025/029	AGTG	NI/NK/NG/NN	βγδε
TAL011/018/025/030	AGTT	NI/NK/NG/NG	βγδε
TAL011/019/021/026	AGAA	NI/NN/NI/NI	βγδε
TAL011/019/021/027	AGAC	NI/NN/NI/HD	βγδε
TAL011/019/021/028	AGAG	NI/NN/NI/NK	βγδε
TAL011/019/021/029	AGAG	NI/NN/NI/NN	βγδε
TAL011/019/021/030	AGAT	NI/NN/NI/NG	βγδε
TAL011/019/022/026	AGCA	NI/NN/HD/NI	βγδε
TAL011/019/022/027	AGCC	NI/NN/HD/HD	βγδε
TAL011/019/022/028	AGCG	NI/NN/HD/NK	βγδε
TAL011/019/022/029	AGCG	NI/NN/HD/NN	βγδε
TAL011/019/022/030	AGCT	NI/NN/HD/NG	βγδε
TAL011/019/023/026	AGGA	NI/NN/NK/NI	βγδε
TAL011/019/023/027	AGGC	NI/NN/NK/HD	βγδε
TAL011/019/023/028	AGGG	NI/NN/NK/NK	βγδε
TAL011/019/023/029	AGGG	NI/NN/NK/NN	βγδε
TAL011/019/023/030	AGGT	NI/NN/NK/NG	βγδε
TAL011/019/024/026	AGGA	NI/NN/NN/NI	βγδε
TAL011/019/024/027	AGGC	NI/NN/NN/HD	βγδε
TAL011/019/024/028	AGGG	NI/NN/NN/NK	βγδε
TAL011/019/024/029	AGGG	NI/NN/NN/NN	βγδε
TAL011/019/024/030	AGGT	NI/NN/NN/NG	βγδε
TAL011/019/025/026	AGTA	NI/NN/NG/NI	βγδε
TAL011/019/025/027	AGTC	NI/NN/NG/HD	βγδε
TAL011/019/025/028	AGTG	NI/NN/NG/NK	βγδε
TAL011/019/025/029	AGTG	NI/NN/NG/NN	βγδε
TAL011/019/025/030	AGTT	NI/NN/NG/NG	βγδε
TAL011/020/021/026	ATAA	NI/NG/NI/NI	βγδε
TAL011/020/021/027	ATAC	NI/NG/NI/HD	βγδε
TAL011/020/021/028	ATAG	NI/NG/NI/NK	βγδε

TAL011/020/021/029	ATAG	NI/NG/NI/NN	βγδε
TAL011/020/021/030	ATAT	NI/NG/NI/NG	βγδε
TAL011/020/022/026	ATCA	NI/NG/HD/NI	βγδε
TAL011/020/022/027	ATCC	NI/NG/HD/HD	βγδε
TAL011/020/022/028	ATCG	NI/NG/HD/NK	βγδε
TAL011/020/022/029	ATCG	NI/NG/HD/NN	βγδε
TAL011/020/022/030	ATCT	NI/NG/HD/NG	βγδε
TAL011/020/023/026	ATGA	NI/NG/NK/NI	βγδε
TAL011/020/023/027	ATGC	NI/NG/NK/HD	βγδε
TAL011/020/023/028	ATGG	NI/NG/NK/NK	βγδε
TAL011/020/023/029	ATGG	NI/NG/NK/NN	βγδε
TAL011/020/023/030	ATGT	NI/NG/NK/NG	βγδε
TAL011/020/024/026	ATGA	NI/NG/NN/NI	βγδε
TAL011/020/024/027	ATGC	NI/NG/NN/HD	βγδε
TAL011/020/024/028	ATGG	NI/NG/NN/NK	βγδε
TAL011/020/024/029	ATGG	NI/NG/NN/NN	βγδε
TAL011/020/024/030	ATGT	NI/NG/NN/NG	βγδε
TAL011/020/025/026	ATTA	NI/NG/NG/NI	βγδε
TAL011/020/025/027	ATTC	NI/NG/NG/HD	βγδε
TAL011/020/025/028	ATTG	NI/NG/NG/NK	βγδε
TAL011/020/025/029	ATTG	NI/NG/NG/NN	βγδε
TAL011/020/025/030	ATTT	NI/NG/NG/NG	βγδε
TAL012/016/021/026	CAAA	HD/NI/NI/NI	βγδε
TAL012/016/021/027	CAAC	HD/NI/NI/HD	βγδε
TAL012/016/021/028	CAAG	HD/NI/NI/NK	βγδε
TAL012/016/021/029	CAAG	HD/NI/NI/NN	βγδε
TAL012/016/021/030	CAAT	HD/NI/NI/NG	βγδε
TAL012/016/022/026	CACA	HD/NI/HD/NI	βγδε
TAL012/016/022/027	CACC	HD/NI/HD/HD	βγδε
TAL012/016/022/028	CACG	HD/NI/HD/NK	βγδε
TAL012/016/022/029	CACG	HD/NI/HD/NN	βγδε
TAL012/016/022/030	CACT	HD/NI/HD/NG	βγδε
TAL012/016/023/026	CAGA	HD/NI/NK/NI	βγδε
TAL012/016/023/027	CAGC	HD/NI/NK/HD	βγδε
TAL012/016/023/028	CAGG	HD/NI/NK/NK	βγδε
TAL012/016/023/029	CAGG	HD/NI/NK/NN	βγδε
TAL012/016/023/030	CAGT	HD/NI/NK/NG	βγδε
TAL012/016/024/026	CAGA	HD/NI/NN/NI	βγδε
TAL012/016/024/027	CAGC	HD/NI/NN/HD	βγδε
TAL012/016/024/028	CAGG	HD/NI/NN/NK	βγδε
TAL012/016/024/029	CAGG	HD/NI/NN/NN	βγδε

TAL012/016/024/030	CAGT	HD/NI/NN/NG	βγδε
TAL012/016/025/026	CATA	HD/NI/NG/NI	βγδε
TAL012/016/025/027	CATC	HD/NI/NG/HD	βγδε
TAL012/016/025/028	CATG	HD/NI/NG/NK	βγδε
TAL012/016/025/029	CATG	HD/NI/NG/NN	βγδε
TAL012/016/025/030	CATT	HD/NI/NG/NG	βγδε
TAL012/017/021/026	CCAA	HD/HD/NI/NI	βγδε
TAL012/017/021/027	CCAC	HD/HD/NI/HD	βγδε
TAL012/017/021/028	CCAG	HD/HD/NI/NK	βγδε
TAL012/017/021/029	CCAG	HD/HD/NI/NN	βγδε
TAL012/017/021/030	CCAT	HD/HD/NI/NG	βγδε
TAL012/017/022/026	CCCA	HD/HD/HD/NI	βγδε
TAL012/017/022/027	CCCC	HD/HD/HD/HD	βγδε
TAL012/017/022/028	CCCG	HD/HD/HD/NK	βγδε
TAL012/017/022/029	CCCG	HD/HD/HD/NN	βγδε
TAL012/017/022/030	CCCT	HD/HD/HD/NG	βγδε
TAL012/017/023/026	CCGA	HD/HD/NK/NI	βγδε
TAL012/017/023/027	CCGC	HD/HD/NK/HD	βγδε
TAL012/017/023/028	CCGG	HD/HD/NK/NK	βγδε
TAL012/017/023/029	CCGG	HD/HD/NK/NN	βγδε
TAL012/017/023/030	CCGT	HD/HD/NK/NG	βγδε
TAL012/017/024/026	CCGA	HD/HD/NN/NI	βγδε
TAL012/017/024/027	CCGC	HD/HD/NN/HD	βγδε
TAL012/017/024/028	CCGG	HD/HD/NN/NK	βγδε
TAL012/017/024/029	CCGG	HD/HD/NN/NN	βγδε
TAL012/017/024/030	CCGT	HD/HD/NN/NG	βγδε
TAL012/017/025/026	CCTA	HD/HD/NG/NI	βγδε
TAL012/017/025/027	CCTC	HD/HD/NG/HD	βγδε
TAL012/017/025/028	CCTG	HD/HD/NG/NK	βγδε
TAL012/017/025/029	CCTG	HD/HD/NG/NN	βγδε
TAL012/017/025/030	CCTT	HD/HD/NG/NG	βγδε
TAL012/018/021/026	CGAA	HD/NK/NI/NI	βγδε
TAL012/018/021/027	CGAC	HD/NK/NI/HD	βγδε
TAL012/018/021/028	CGAG	HD/NK/NI/NK	βγδε
TAL012/018/021/029	CGAG	HD/NK/NI/NN	βγδε
TAL012/018/021/030	CGAT	HD/NK/NI/NG	βγδε
TAL012/018/022/026	CGCA	HD/NK/HD/NI	βγδε
TAL012/018/022/027	CGCC	HD/NK/HD/HD	βγδε
TAL012/018/022/028	CGCG	HD/NK/HD/NK	βγδε
TAL012/018/022/029	CGCG	HD/NK/HD/NN	βγδε
TAL012/018/022/030	CGCT	HD/NK/HD/NG	βγδε

TAL012/018/023/026	CGGA	HD/NK/NK/NI	βγδε
TAL012/018/023/027	CGGC	HD/NK/NK/HD	βγδε
TAL012/018/023/028	CGGG	HD/NK/NK/NK	βγδε
TAL012/018/023/029	CGGG	HD/NK/NK/NN	βγδε
TAL012/018/023/030	CGGT	HD/NK/NK/NG	βγδε
TAL012/018/024/026	CGGA	HD/NK/NN/NI	βγδε
TAL012/018/024/027	CGGC	HD/NK/NN/HD	βγδε
TAL012/018/024/028	CGGG	HD/NK/NN/NK	βγδε
TAL012/018/024/029	CGGG	HD/NK/NN/NN	βγδε
TAL012/018/024/030	CGGT	HD/NK/NN/NG	βγδε
TAL012/018/025/026	CGTA	HD/NK/NG/NI	βγδε
TAL012/018/025/027	CGTC	HD/NK/NG/HD	βγδε
TAL012/018/025/028	CGTG	HD/NK/NG/NK	βγδε
TAL012/018/025/029	CGTG	HD/NK/NG/NN	βγδε
TAL012/018/025/030	CGTT	HD/NK/NG/NG	βγδε
TAL012/019/021/026	CGAA	HD/NN/NI/NI	βγδε
TAL012/019/021/027	CGAC	HD/NN/NI/HD	βγδε
TAL012/019/021/028	CGAG	HD/NN/NI/NK	βγδε
TAL012/019/021/029	CGAG	HD/NN/NI/NN	βγδε
TAL012/019/021/030	CGAT	HD/NN/NI/NG	βγδε
TAL012/019/022/026	CGCA	HD/NN/HD/NI	βγδε
TAL012/019/022/027	CGCC	HD/NN/HD/HD	βγδε
TAL012/019/022/028	CGCG	HD/NN/HD/NK	βγδε
TAL012/019/022/029	CGCG	HD/NN/HD/NN	βγδε
TAL012/019/022/030	CGCT	HD/NN/HD/NG	βγδε
TAL012/019/023/026	CGGA	HD/NN/NK/NI	βγδε
TAL012/019/023/027	CGGC	HD/NN/NK/HD	βγδε
TAL012/019/023/028	CGGG	HD/NN/NK/NK	βγδε
TAL012/019/023/029	CGGG	HD/NN/NK/NN	βγδε
TAL012/019/023/030	CGGT	HD/NN/NK/NG	βγδε
TAL012/019/024/026	CGGA	HD/NN/NN/NI	βγδε
TAL012/019/024/027	CGGC	HD/NN/NN/HD	βγδε
TAL012/019/024/028	CGGG	HD/NN/NN/NK	βγδε
TAL012/019/024/029	CGGG	HD/NN/NN/NN	βγδε
TAL012/019/024/030	CGGT	HD/NN/NN/NG	βγδε
TAL012/019/025/026	CGTA	HD/NN/NG/NI	βγδε
TAL012/019/025/027	CGTC	HD/NN/NG/HD	βγδε
TAL012/019/025/028	CGTG	HD/NN/NG/NK	βγδε
TAL012/019/025/029	CGTG	HD/NN/NG/NN	βγδε
TAL012/019/025/030	CGTT	HD/NN/NG/NG	βγδε
TAL012/020/021/026	CTAA	HD/NG/NI/NI	βγδε

TAL012/020/021/027	CTAC	HD/NG/NI/HD	βγδε
TAL012/020/021/028	CTAG	HD/NG/NI/NK	βγδε
TAL012/020/021/029	CTAG	HD/NG/NI/NN	βγδε
TAL012/020/021/030	CTAT	HD/NG/NI/NG	βγδε
TAL012/020/022/026	CTCA	HD/NG/HD/NI	βγδε
TAL012/020/022/027	CTCC	HD/NG/HD/HD	βγδε
TAL012/020/022/028	CTCG	HD/NG/HD/NK	βγδε
TAL012/020/022/029	CTCG	HD/NG/HD/NN	βγδε
TAL012/020/022/030	CTCT	HD/NG/HD/NG	βγδε
TAL012/020/023/026	CTGA	HD/NG/NK/NI	βγδε
TAL012/020/023/027	CTGC	HD/NG/NK/HD	βγδε
TAL012/020/023/028	CTGG	HD/NG/NK/NK	βγδε
TAL012/020/023/029	CTGG	HD/NG/NK/NN	βγδε
TAL012/020/023/030	CTGT	HD/NG/NK/NG	βγδε
TAL012/020/024/026	CTGA	HD/NG/NN/NI	βγδε
TAL012/020/024/027	CTGC	HD/NG/NN/HD	βγδε
TAL012/020/024/028	CTGG	HD/NG/NN/NK	βγδε
TAL012/020/024/029	CTGG	HD/NG/NN/NN	βγδε
TAL012/020/024/030	CTGT	HD/NG/NN/NG	βγδε
TAL012/020/025/026	CTTA	HD/NG/NG/NI	βγδε
TAL012/020/025/027	CTTC	HD/NG/NG/HD	βγδε
TAL012/020/025/028	CTTG	HD/NG/NG/NK	βγδε
TAL012/020/025/029	CTTG	HD/NG/NG/NN	βγδε
TAL012/020/025/030	CTTT	HD/NG/NG/NG	βγδε
TAL013/016/021/026	GAAA	NK/NI/NI/NI	βγδε
TAL013/016/021/027	GAAC	NK/NI/NI/HD	βγδε
TAL013/016/021/028	GAAG	NK/NI/NI/NK	βγδε
TAL013/016/021/029	GAAG	NK/NI/NI/NN	βγδε
TAL013/016/021/030	GAAT	NK/NI/NI/NG	βγδε
TAL013/016/022/026	GACA	NK/NI/HD/NI	βγδε
TAL013/016/022/027	GACC	NK/NI/HD/HD	βγδε
TAL013/016/022/028	GACG	NK/NI/HD/NK	βγδε
TAL013/016/022/029	GACG	NK/NI/HD/NN	βγδε
TAL013/016/022/030	GACT	NK/NI/HD/NG	βγδε
TAL013/016/023/026	GAGA	NK/NI/NK/NI	βγδε
TAL013/016/023/027	GAGC	NK/NI/NK/HD	βγδε
TAL013/016/023/028	GAGG	NK/NI/NK/NK	βγδε
TAL013/016/023/029	GAGG	NK/NI/NK/NN	βγδε
TAL013/016/023/030	GAGT	NK/NI/NK/NG	βγδε
TAL013/016/024/026	GAGA	NK/NI/NN/NI	βγδε
TAL013/016/024/027	GAGC	NK/NI/NN/HD	βγδε

TAL013/016/024/028	GAGG	NK/NI/NN/NK	$\beta\gamma\delta\epsilon$
TAL013/016/024/029	GAGG	NK/NI/NN/NN	$\beta\gamma\delta\epsilon$
TAL013/016/024/030	GAGT	NK/NI/NN/NG	$\beta\gamma\delta\epsilon$
TAL013/016/025/026	GATA	NK/NI/NG/NI	$\beta\gamma\delta\epsilon$
TAL013/016/025/027	GATC	NK/NI/NG/HD	$\beta\gamma\delta\epsilon$
TAL013/016/025/028	GATG	NK/NI/NG/NK	$\beta\gamma\delta\epsilon$
TAL013/016/025/029	GATG	NK/NI/NG/NN	$\beta\gamma\delta\epsilon$
TAL013/016/025/030	GATT	NK/NI/NG/NG	$\beta\gamma\delta\epsilon$
TAL013/017/021/026	GCAA	NK/HD/NI/NI	$\beta\gamma\delta\epsilon$
TAL013/017/021/027	GCAC	NK/HD/NI/HD	$\beta\gamma\delta\epsilon$
TAL013/017/021/028	GCAG	NK/HD/NI/NK	$\beta\gamma\delta\epsilon$
TAL013/017/021/029	GCAG	NK/HD/NI/NN	$\beta\gamma\delta\epsilon$
TAL013/017/021/030	GCAT	NK/HD/NI/NG	$\beta\gamma\delta\epsilon$
TAL013/017/022/026	GCCA	NK/HD/HD/NI	$\beta\gamma\delta\epsilon$
TAL013/017/022/027	GCCC	NK/HD/HD/HD	$\beta\gamma\delta\epsilon$
TAL013/017/022/028	GCCG	NK/HD/HD/NK	$\beta\gamma\delta\epsilon$
TAL013/017/022/029	GCCG	NK/HD/HD/NN	$\beta\gamma\delta\epsilon$
TAL013/017/022/030	GCCT	NK/HD/HD/NG	$\beta\gamma\delta\epsilon$
TAL013/017/023/026	GCGA	NK/HD/NK/NI	$\beta\gamma\delta\epsilon$
TAL013/017/023/027	GCGC	NK/HD/NK/HD	$\beta\gamma\delta\epsilon$
TAL013/017/023/028	GCGG	NK/HD/NK/NK	$\beta\gamma\delta\epsilon$
TAL013/017/023/029	GCGG	NK/HD/NK/NN	$\beta\gamma\delta\epsilon$
TAL013/017/023/030	GCGT	NK/HD/NK/NG	$\beta\gamma\delta\epsilon$
TAL013/017/024/026	GCGA	NK/HD/NN/NI	$\beta\gamma\delta\epsilon$
TAL013/017/024/027	GCGC	NK/HD/NN/HD	$\beta\gamma\delta\epsilon$
TAL013/017/024/028	GCGG	NK/HD/NN/NK	$\beta\gamma\delta\epsilon$
TAL013/017/024/029	GCGG	NK/HD/NN/NN	$\beta\gamma\delta\epsilon$
TAL013/017/024/030	GCGT	NK/HD/NN/NG	$\beta\gamma\delta\epsilon$
TAL013/017/025/026	GCTA	NK/HD/NG/NI	$\beta\gamma\delta\epsilon$
TAL013/017/025/027	GCTC	NK/HD/NG/HD	$\beta\gamma\delta\epsilon$
TAL013/017/025/028	GCTG	NK/HD/NG/NK	$\beta\gamma\delta\epsilon$
TAL013/017/025/029	GCTG	NK/HD/NG/NN	$\beta\gamma\delta\epsilon$
TAL013/017/025/030	GCTT	NK/HD/NG/NG	$\beta\gamma\delta\epsilon$
TAL013/018/021/026	GGAA	NK/NK/NI/NI	$\beta\gamma\delta\epsilon$
TAL013/018/021/027	GGAC	NK/NK/NI/HD	$\beta\gamma\delta\epsilon$
TAL013/018/021/028	GGAG	NK/NK/NI/NK	$\beta\gamma\delta\epsilon$
TAL013/018/021/029	GGAG	NK/NK/NI/NN	$\beta\gamma\delta\epsilon$
TAL013/018/021/030	GGAT	NK/NK/NI/NG	$\beta\gamma\delta\epsilon$
TAL013/018/022/026	GGCA	NK/NK/HD/NI	$\beta\gamma\delta\epsilon$
TAL013/018/022/027	GGCC	NK/NK/HD/HD	$\beta\gamma\delta\epsilon$
TAL013/018/022/028	GGCG	NK/NK/HD/NK	$\beta\gamma\delta\epsilon$

TAL013/018/022/029	GGCG	NK/NK/HD/NN	βγδε
TAL013/018/022/030	GGCT	NK/NK/HD/NG	βγδε
TAL013/018/023/026	GGGA	NK/NK/NK/NI	βγδε
TAL013/018/023/027	GGGC	NK/NK/NK/HD	βγδε
TAL013/018/023/028	GGGG	NK/NK/NK/NK	βγδε
TAL013/018/023/029	GGGG	NK/NK/NK/NN	βγδε
TAL013/018/023/030	GGGT	NK/NK/NK/NG	βγδε
TAL013/018/024/026	GGGA	NK/NK/NN/NI	βγδε
TAL013/018/024/027	GGGC	NK/NK/NN/HD	βγδε
TAL013/018/024/028	GGGG	NK/NK/NN/NK	βγδε
TAL013/018/024/029	GGGG	NK/NK/NN/NN	βγδε
TAL013/018/024/030	GGGT	NK/NK/NN/NG	βγδε
TAL013/018/025/026	GGTA	NK/NK/NG/NI	βγδε
TAL013/018/025/027	GGTC	NK/NK/NG/HD	βγδε
TAL013/018/025/028	GGTG	NK/NK/NG/NK	βγδε
TAL013/018/025/029	GGTG	NK/NK/NG/NN	βγδε
TAL013/018/025/030	GGTT	NK/NK/NG/NG	βγδε
TAL013/019/021/026	GGAA	NK/NN/NI/NI	βγδε
TAL013/019/021/027	GGAC	NK/NN/NI/HD	βγδε
TAL013/019/021/028	GGAG	NK/NN/NI/NK	βγδε
TAL013/019/021/029	GGAG	NK/NN/NI/NN	βγδε
TAL013/019/021/030	GGAT	NK/NN/NI/NG	βγδε
TAL013/019/022/026	GGCA	NK/NN/HD/NI	βγδε
TAL013/019/022/027	GGCC	NK/NN/HD/HD	βγδε
TAL013/019/022/028	GGCG	NK/NN/HD/NK	βγδε
TAL013/019/022/029	GGCG	NK/NN/HD/NN	βγδε
TAL013/019/022/030	GGCT	NK/NN/HD/NG	βγδε
TAL013/019/023/026	GGGA	NK/NN/NK/NI	βγδε
TAL013/019/023/027	GGGC	NK/NN/NK/HD	βγδε
TAL013/019/023/028	GGGG	NK/NN/NK/NK	βγδε
TAL013/019/023/029	GGGG	NK/NN/NK/NN	βγδε
TAL013/019/023/030	GGGT	NK/NN/NK/NG	βγδε
TAL013/019/024/026	GGGA	NK/NN/NN/NI	βγδε
TAL013/019/024/027	GGGC	NK/NN/NN/HD	βγδε
TAL013/019/024/028	GGGG	NK/NN/NN/NK	βγδε
TAL013/019/024/029	GGGG	NK/NN/NN/NN	βγδε
TAL013/019/024/030	GGGT	NK/NN/NN/NG	βγδε
TAL013/019/025/026	GGTA	NK/NN/NG/NI	βγδε
TAL013/019/025/027	GGTC	NK/NN/NG/HD	βγδε
TAL013/019/025/028	GGTG	NK/NN/NG/NK	βγδε
TAL013/019/025/029	GGTG	NK/NN/NG/NN	βγδε

TAL013/019/025/030	GGTT	NK/NN/NG/NG	βγδε
TAL013/020/021/026	GTAA	NK/NG/NI/NI	βγδε
TAL013/020/021/027	GTAC	NK/NG/NI/HD	βγδε
TAL013/020/021/028	GTAG	NK/NG/NI/NK	βγδε
TAL013/020/021/029	GTAG	NK/NG/NI/NN	βγδε
TAL013/020/021/030	GTAT	NK/NG/NI/NG	βγδε
TAL013/020/022/026	GTCA	NK/NG/HD/NI	βγδε
TAL013/020/022/027	GTCC	NK/NG/HD/HD	βγδε
TAL013/020/022/028	GTCG	NK/NG/HD/NK	βγδε
TAL013/020/022/029	GTCG	NK/NG/HD/NN	βγδε
TAL013/020/022/030	GTCT	NK/NG/HD/NG	βγδε
TAL013/020/023/026	GTGA	NK/NG/NK/NI	βγδε
TAL013/020/023/027	GTGC	NK/NG/NK/HD	βγδε
TAL013/020/023/028	GTGG	NK/NG/NK/NK	βγδε
TAL013/020/023/029	GTGG	NK/NG/NK/NN	βγδε
TAL013/020/023/030	GTGT	NK/NG/NK/NG	βγδε
TAL013/020/024/026	GTGA	NK/NG/NN/NI	βγδε
TAL013/020/024/027	GTGC	NK/NG/NN/HD	βγδε
TAL013/020/024/028	GTGG	NK/NG/NN/NK	βγδε
TAL013/020/024/029	GTGG	NK/NG/NN/NN	βγδε
TAL013/020/024/030	GTGT	NK/NG/NN/NG	βγδε
TAL013/020/025/026	GTTA	NK/NG/NG/NI	βγδε
TAL013/020/025/027	G TTC	NK/NG/NG/HD	βγδε
TAL013/020/025/028	GTTG	NK/NG/NG/NK	βγδε
TAL013/020/025/029	GTTG	NK/NG/NG/NN	βγδε
TAL013/020/025/030	GTTT	NK/NG/NG/NG	βγδε
TAL014/016/021/026	GAAA	NN/NI/NI/NI	βγδε
TAL014/016/021/027	GAAC	NN/NI/NI/HD	βγδε
TAL014/016/021/028	GAAG	NN/NI/NI/NK	βγδε
TAL014/016/021/029	GAAG	NN/NI/NI/NN	βγδε
TAL014/016/021/030	GAAT	NN/NI/NI/NG	βγδε
TAL014/016/022/026	GACA	NN/NI/HD/NI	βγδε
TAL014/016/022/027	GACC	NN/NI/HD/HD	βγδε
TAL014/016/022/028	GACG	NN/NI/HD/NK	βγδε
TAL014/016/022/029	GACG	NN/NI/HD/NN	βγδε
TAL014/016/022/030	GACT	NN/NI/HD/NG	βγδε
TAL014/016/023/026	GAGA	NN/NI/NK/NI	βγδε
TAL014/016/023/027	GAGC	NN/NI/NK/HD	βγδε
TAL014/016/023/028	GAGG	NN/NI/NK/NK	βγδε
TAL014/016/023/029	GAGG	NN/NI/NK/NN	βγδε
TAL014/016/023/030	GAGT	NN/NI/NK/NG	βγδε

TAL014/016/024/026	GAGA	NN/NI/NN/NI	βγδε
TAL014/016/024/027	GAGC	NN/NI/NN/HD	βγδε
TAL014/016/024/028	GAGG	NN/NI/NN/NK	βγδε
TAL014/016/024/029	GAGG	NN/NI/NN/NN	βγδε
TAL014/016/024/030	GAGT	NN/NI/NN/NG	βγδε
TAL014/016/025/026	GATA	NN/NI/NG/NI	βγδε
TAL014/016/025/027	GATC	NN/NI/NG/HD	βγδε
TAL014/016/025/028	GATG	NN/NI/NG/NK	βγδε
TAL014/016/025/029	GATG	NN/NI/NG/NN	βγδε
TAL014/016/025/030	GATT	NN/NI/NG/NG	βγδε
TAL014/017/021/026	GCAA	NN/HD/NI/NI	βγδε
TAL014/017/021/027	GCAC	NN/HD/NI/HD	βγδε
TAL014/017/021/028	GCAG	NN/HD/NI/NK	βγδε
TAL014/017/021/029	GCAG	NN/HD/NI/NN	βγδε
TAL014/017/021/030	GCAT	NN/HD/NI/NG	βγδε
TAL014/017/022/026	GCCA	NN/HD/HD/NI	βγδε
TAL014/017/022/027	GCCC	NN/HD/HD/HD	βγδε
TAL014/017/022/028	GCCG	NN/HD/HD/NK	βγδε
TAL014/017/022/029	GCCG	NN/HD/HD/NN	βγδε
TAL014/017/022/030	GCCT	NN/HD/HD/NG	βγδε
TAL014/017/023/026	GCGA	NN/HD/NK/NI	βγδε
TAL014/017/023/027	GCGC	NN/HD/NK/HD	βγδε
TAL014/017/023/028	GCGG	NN/HD/NK/NK	βγδε
TAL014/017/023/029	GCGG	NN/HD/NK/NN	βγδε
TAL014/017/023/030	GCGT	NN/HD/NK/NG	βγδε
TAL014/017/024/026	GCGA	NN/HD/NN/NI	βγδε
TAL014/017/024/027	GCGC	NN/HD/NN/HD	βγδε
TAL014/017/024/028	GCGG	NN/HD/NN/NK	βγδε
TAL014/017/024/029	GCGG	NN/HD/NN/NN	βγδε
TAL014/017/024/030	GCGT	NN/HD/NN/NG	βγδε
TAL014/017/025/026	GCTA	NN/HD/NG/NI	βγδε
TAL014/017/025/027	GCTC	NN/HD/NG/HD	βγδε
TAL014/017/025/028	GCTG	NN/HD/NG/NK	βγδε
TAL014/017/025/029	GCTG	NN/HD/NG/NN	βγδε
TAL014/017/025/030	GCTT	NN/HD/NG/NG	βγδε
TAL014/018/021/026	GGAA	NN/NK/NI/NI	βγδε
TAL014/018/021/027	GGAC	NN/NK/NI/HD	βγδε
TAL014/018/021/028	GGAG	NN/NK/NI/NK	βγδε
TAL014/018/021/029	GGAG	NN/NK/NI/NN	βγδε
TAL014/018/021/030	GGAT	NN/NK/NI/NG	βγδε
TAL014/018/022/026	GGCA	NN/NK/HD/NI	βγδε

TAL014/018/022/027	GGCC	NN/NK/HD/HD	βγδε
TAL014/018/022/028	GGCG	NN/NK/HD/NK	βγδε
TAL014/018/022/029	GGCG	NN/NK/HD/NN	βγδε
TAL014/018/022/030	GGCT	NN/NK/HD/NG	βγδε
TAL014/018/023/026	GGGA	NN/NK/NK/NI	βγδε
TAL014/018/023/027	GGGC	NN/NK/NK/HD	βγδε
TAL014/018/023/028	GGGG	NN/NK/NK/NK	βγδε
TAL014/018/023/029	GGGG	NN/NK/NK/NN	βγδε
TAL014/018/023/030	GGGT	NN/NK/NK/NG	βγδε
TAL014/018/024/026	GGGA	NN/NK/NN/NI	βγδε
TAL014/018/024/027	GGGC	NN/NK/NN/HD	βγδε
TAL014/018/024/028	GGGG	NN/NK/NN/NK	βγδε
TAL014/018/024/029	GGGG	NN/NK/NN/NN	βγδε
TAL014/018/024/030	GGGT	NN/NK/NN/NG	βγδε
TAL014/018/025/026	GGTA	NN/NK/NG/NI	βγδε
TAL014/018/025/027	GGTC	NN/NK/NG/HD	βγδε
TAL014/018/025/028	GGTG	NN/NK/NG/NK	βγδε
TAL014/018/025/029	GGTG	NN/NK/NG/NN	βγδε
TAL014/018/025/030	GGTT	NN/NK/NG/NG	βγδε
TAL014/019/021/026	GGAA	NN/NN/NI/NI	βγδε
TAL014/019/021/027	GGAC	NN/NN/NI/HD	βγδε
TAL014/019/021/028	GGAG	NN/NN/NI/NK	βγδε
TAL014/019/021/029	GGAG	NN/NN/NI/NN	βγδε
TAL014/019/021/030	GGAT	NN/NN/NI/NG	βγδε
TAL014/019/022/026	GGCA	NN/NN/HD/NI	βγδε
TAL014/019/022/027	GGCC	NN/NN/HD/HD	βγδε
TAL014/019/022/028	GGCG	NN/NN/HD/NK	βγδε
TAL014/019/022/029	GGCG	NN/NN/HD/NN	βγδε
TAL014/019/022/030	GGCT	NN/NN/HD/NG	βγδε
TAL014/019/023/026	GGGA	NN/NN/NK/NI	βγδε
TAL014/019/023/027	GGGC	NN/NN/NK/HD	βγδε
TAL014/019/023/028	GGGG	NN/NN/NK/NK	βγδε
TAL014/019/023/029	GGGG	NN/NN/NK/NN	βγδε
TAL014/019/023/030	GGGT	NN/NN/NK/NG	βγδε
TAL014/019/024/026	GGGA	NN/NN/NN/NI	βγδε
TAL014/019/024/027	GGGC	NN/NN/NN/HD	βγδε
TAL014/019/024/028	GGGG	NN/NN/NN/NK	βγδε
TAL014/019/024/029	GGGG	NN/NN/NN/NN	βγδε
TAL014/019/024/030	GGGT	NN/NN/NN/NG	βγδε
TAL014/019/025/026	GGTA	NN/NN/NG/NI	βγδε
TAL014/019/025/027	GGTC	NN/NN/NG/HD	βγδε

TAL014/019/025/028	GGTG	NN/NN/NG/NK	βγδε
TAL014/019/025/029	GGTG	NN/NN/NG/NN	βγδε
TAL014/019/025/030	GGTT	NN/NN/NG/NG	βγδε
TAL014/020/021/026	GTAA	NN/NG/NI/NI	βγδε
TAL014/020/021/027	GTAC	NN/NG/NI/HD	βγδε
TAL014/020/021/028	GTAG	NN/NG/NI/NK	βγδε
TAL014/020/021/029	GTAG	NN/NG/NI/NN	βγδε
TAL014/020/021/030	GTAT	NN/NG/NI/NG	βγδε
TAL014/020/022/026	GTCA	NN/NG/HD/NI	βγδε
TAL014/020/022/027	GTCC	NN/NG/HD/HD	βγδε
TAL014/020/022/028	GTCG	NN/NG/HD/NK	βγδε
TAL014/020/022/029	GTCG	NN/NG/HD/NN	βγδε
TAL014/020/022/030	GTCT	NN/NG/HD/NG	βγδε
TAL014/020/023/026	GTGA	NN/NG/NK/NI	βγδε
TAL014/020/023/027	GTGC	NN/NG/NK/HD	βγδε
TAL014/020/023/028	GTGG	NN/NG/NK/NK	βγδε
TAL014/020/023/029	GTGG	NN/NG/NK/NN	βγδε
TAL014/020/023/030	GTGT	NN/NG/NK/NG	βγδε
TAL014/020/024/026	GTGA	NN/NG/NN/NI	βγδε
TAL014/020/024/027	GTGC	NN/NG/NN/HD	βγδε
TAL014/020/024/028	GTGG	NN/NG/NN/NK	βγδε
TAL014/020/024/029	GTGG	NN/NG/NN/NN	βγδε
TAL014/020/024/030	GTGT	NN/NG/NN/NG	βγδε
TAL014/020/025/026	GTTA	NN/NG/NG/NI	βγδε
TAL014/020/025/027	GTTC	NN/NG/NG/HD	βγδε
TAL014/020/025/028	GTTG	NN/NG/NG/NK	βγδε
TAL014/020/025/029	GTTG	NN/NG/NG/NN	βγδε
TAL014/020/025/030	GTTT	NN/NG/NG/NG	βγδε
TAL015/016/021/026	TAAA	NG/NI/NI/NI	βγδε
TAL015/016/021/027	TAAC	NG/NI/NI/HD	βγδε
TAL015/016/021/028	TAAG	NG/NI/NI/NK	βγδε
TAL015/016/021/029	TAAG	NG/NI/NI/NN	βγδε
TAL015/016/021/030	TAAT	NG/NI/NI/NG	βγδε
TAL015/016/022/026	TACA	NG/NI/HD/NI	βγδε
TAL015/016/022/027	TACC	NG/NI/HD/HD	βγδε
TAL015/016/022/028	TACG	NG/NI/HD/NK	βγδε
TAL015/016/022/029	TACG	NG/NI/HD/NN	βγδε
TAL015/016/022/030	TACT	NG/NI/HD/NG	βγδε
TAL015/016/023/026	TAGA	NG/NI/NK/NI	βγδε
TAL015/016/023/027	TAGC	NG/NI/NK/HD	βγδε
TAL015/016/023/028	TAGG	NG/NI/NK/NK	βγδε

TAL015/016/023/029	TAGG	NG/NI/NK/NN	$\beta\gamma\delta\epsilon$
TAL015/016/023/030	TAGT	NG/NI/NK/NG	$\beta\gamma\delta\epsilon$
TAL015/016/024/026	TAGA	NG/NI/NN/NI	$\beta\gamma\delta\epsilon$
TAL015/016/024/027	TAGC	NG/NI/NN/HD	$\beta\gamma\delta\epsilon$
TAL015/016/024/028	TAGG	NG/NI/NN/NK	$\beta\gamma\delta\epsilon$
TAL015/016/024/029	TAGG	NG/NI/NN/NN	$\beta\gamma\delta\epsilon$
TAL015/016/024/030	TAGT	NG/NI/NN/NG	$\beta\gamma\delta\epsilon$
TAL015/016/025/026	TATA	NG/NI/NG/NI	$\beta\gamma\delta\epsilon$
TAL015/016/025/027	TATC	NG/NI/NG/HD	$\beta\gamma\delta\epsilon$
TAL015/016/025/028	TATG	NG/NI/NG/NK	$\beta\gamma\delta\epsilon$
TAL015/016/025/029	TATG	NG/NI/NG/NN	$\beta\gamma\delta\epsilon$
TAL015/016/025/030	TATT	NG/NI/NG/NG	$\beta\gamma\delta\epsilon$
TAL015/017/021/026	TCAA	NG/HD/NI/NI	$\beta\gamma\delta\epsilon$
TAL015/017/021/027	TCAC	NG/HD/NI/HD	$\beta\gamma\delta\epsilon$
TAL015/017/021/028	TCAG	NG/HD/NI/NK	$\beta\gamma\delta\epsilon$
TAL015/017/021/029	TCAG	NG/HD/NI/NN	$\beta\gamma\delta\epsilon$
TAL015/017/021/030	TCAT	NG/HD/NI/NG	$\beta\gamma\delta\epsilon$
TAL015/017/022/026	TCCA	NG/HD/HD/NI	$\beta\gamma\delta\epsilon$
TAL015/017/022/027	TCCC	NG/HD/HD/HD	$\beta\gamma\delta\epsilon$
TAL015/017/022/028	TCCG	NG/HD/HD/NK	$\beta\gamma\delta\epsilon$
TAL015/017/022/029	TCCG	NG/HD/HD/NN	$\beta\gamma\delta\epsilon$
TAL015/017/022/030	TCCT	NG/HD/HD/NG	$\beta\gamma\delta\epsilon$
TAL015/017/023/026	TCGA	NG/HD/NK/NI	$\beta\gamma\delta\epsilon$
TAL015/017/023/027	TCGC	NG/HD/NK/HD	$\beta\gamma\delta\epsilon$
TAL015/017/023/028	TCGG	NG/HD/NK/NK	$\beta\gamma\delta\epsilon$
TAL015/017/023/029	TCGG	NG/HD/NK/NN	$\beta\gamma\delta\epsilon$
TAL015/017/023/030	TCGT	NG/HD/NK/NG	$\beta\gamma\delta\epsilon$
TAL015/017/024/026	TCGA	NG/HD/NN/NI	$\beta\gamma\delta\epsilon$
TAL015/017/024/027	TCGC	NG/HD/NN/HD	$\beta\gamma\delta\epsilon$
TAL015/017/024/028	TCGG	NG/HD/NN/NK	$\beta\gamma\delta\epsilon$
TAL015/017/024/029	TCGG	NG/HD/NN/NN	$\beta\gamma\delta\epsilon$
TAL015/017/024/030	TCGT	NG/HD/NN/NG	$\beta\gamma\delta\epsilon$
TAL015/017/025/026	TCTA	NG/HD/NG/NI	$\beta\gamma\delta\epsilon$
TAL015/017/025/027	TCTC	NG/HD/NG/HD	$\beta\gamma\delta\epsilon$
TAL015/017/025/028	TCTG	NG/HD/NG/NK	$\beta\gamma\delta\epsilon$
TAL015/017/025/029	TCTG	NG/HD/NG/NN	$\beta\gamma\delta\epsilon$
TAL015/017/025/030	TCTT	NG/HD/NG/NG	$\beta\gamma\delta\epsilon$
TAL015/018/021/026	TGAA	NG/NK/NI/NI	$\beta\gamma\delta\epsilon$
TAL015/018/021/027	TGAC	NG/NK/NI/HD	$\beta\gamma\delta\epsilon$
TAL015/018/021/028	TGAG	NG/NK/NI/NK	$\beta\gamma\delta\epsilon$
TAL015/018/021/029	TGAG	NG/NK/NI/NN	$\beta\gamma\delta\epsilon$

TAL015/018/021/030	TGAT	NG/NK/NI/NG	βγδε
TAL015/018/022/026	TGCA	NG/NK/HD/NI	βγδε
TAL015/018/022/027	TGCC	NG/NK/HD/HD	βγδε
TAL015/018/022/028	TGCG	NG/NK/HD/NK	βγδε
TAL015/018/022/029	TGCG	NG/NK/HD/NN	βγδε
TAL015/018/022/030	TGCT	NG/NK/HD/NG	βγδε
TAL015/018/023/026	TGGA	NG/NK/NK/NI	βγδε
TAL015/018/023/027	TGGC	NG/NK/NK/HD	βγδε
TAL015/018/023/028	TGGG	NG/NK/NK/NK	βγδε
TAL015/018/023/029	TGGG	NG/NK/NK/NN	βγδε
TAL015/018/023/030	TGGT	NG/NK/NK/NG	βγδε
TAL015/018/024/026	TGGA	NG/NK/NN/NI	βγδε
TAL015/018/024/027	TGGC	NG/NK/NN/HD	βγδε
TAL015/018/024/028	TGGG	NG/NK/NN/NK	βγδε
TAL015/018/024/029	TGGG	NG/NK/NN/NN	βγδε
TAL015/018/024/030	TGGT	NG/NK/NN/NG	βγδε
TAL015/018/025/026	TGTA	NG/NK/NG/NI	βγδε
TAL015/018/025/027	TGTC	NG/NK/NG/HD	βγδε
TAL015/018/025/028	TGTG	NG/NK/NG/NK	βγδε
TAL015/018/025/029	TGTG	NG/NK/NG/NN	βγδε
TAL015/018/025/030	TGTT	NG/NK/NG/NG	βγδε
TAL015/019/021/026	TGAA	NG/NN/NI/NI	βγδε
TAL015/019/021/027	TGAC	NG/NN/NI/HD	βγδε
TAL015/019/021/028	TGAG	NG/NN/NI/NK	βγδε
TAL015/019/021/029	TGAG	NG/NN/NI/NN	βγδε
TAL015/019/021/030	TGAT	NG/NN/NI/NG	βγδε
TAL015/019/022/026	TGCA	NG/NN/HD/NI	βγδε
TAL015/019/022/027	TGCC	NG/NN/HD/HD	βγδε
TAL015/019/022/028	TGCG	NG/NN/HD/NK	βγδε
TAL015/019/022/029	TGCG	NG/NN/HD/NN	βγδε
TAL015/019/022/030	TGCT	NG/NN/HD/NG	βγδε
TAL015/019/023/026	TGGA	NG/NN/NK/NI	βγδε
TAL015/019/023/027	TGGC	NG/NN/NK/HD	βγδε
TAL015/019/023/028	TGGG	NG/NN/NK/NK	βγδε
TAL015/019/023/029	TGGG	NG/NN/NK/NN	βγδε
TAL015/019/023/030	TGGT	NG/NN/NK/NG	βγδε
TAL015/019/024/026	TGGA	NG/NN/NN/NI	βγδε
TAL015/019/024/027	TGGC	NG/NN/NN/HD	βγδε
TAL015/019/024/028	TGGG	NG/NN/NN/NK	βγδε
TAL015/019/024/029	TGGG	NG/NN/NN/NN	βγδε
TAL015/019/024/030	TGGT	NG/NN/NN/NG	βγδε

TAL015/019/025/026	TGTA	NG/NN/NG/NI	$\beta\gamma\delta\epsilon$
TAL015/019/025/027	TGTC	NG/NN/NG/HD	$\beta\gamma\delta\epsilon$
TAL015/019/025/028	TGTG	NG/NN/NG/NK	$\beta\gamma\delta\epsilon$
TAL015/019/025/029	TGTG	NG/NN/NG/NN	$\beta\gamma\delta\epsilon$
TAL015/019/025/030	TGTT	NG/NN/NG/NG	$\beta\gamma\delta\epsilon$
TAL015/020/021/026	TTAA	NG/NG/NI/NI	$\beta\gamma\delta\epsilon$
TAL015/020/021/027	TTAC	NG/NG/NI/HD	$\beta\gamma\delta\epsilon$
TAL015/020/021/028	TTAG	NG/NG/NI/NK	$\beta\gamma\delta\epsilon$
TAL015/020/021/029	TTAG	NG/NG/NI/NN	$\beta\gamma\delta\epsilon$
TAL015/020/021/030	TTAT	NG/NG/NI/NG	$\beta\gamma\delta\epsilon$
TAL015/020/022/026	TTCA	NG/NG/HD/NI	$\beta\gamma\delta\epsilon$
TAL015/020/022/027	TTCC	NG/NG/HD/HD	$\beta\gamma\delta\epsilon$
TAL015/020/022/028	TTCG	NG/NG/HD/NK	$\beta\gamma\delta\epsilon$
TAL015/020/022/029	TTCG	NG/NG/HD/NN	$\beta\gamma\delta\epsilon$
TAL015/020/022/030	TTCT	NG/NG/HD/NG	$\beta\gamma\delta\epsilon$
TAL015/020/023/026	TTGA	NG/NG/NK/NI	$\beta\gamma\delta\epsilon$
TAL015/020/023/027	TTGC	NG/NG/NK/HD	$\beta\gamma\delta\epsilon$
TAL015/020/023/028	TTGG	NG/NG/NK/NK	$\beta\gamma\delta\epsilon$
TAL015/020/023/029	TTGG	NG/NG/NK/NN	$\beta\gamma\delta\epsilon$
TAL015/020/023/030	TTGT	NG/NG/NK/NG	$\beta\gamma\delta\epsilon$
TAL015/020/024/026	TTGA	NG/NG/NN/NI	$\beta\gamma\delta\epsilon$
TAL015/020/024/027	TTGC	NG/NG/NN/HD	$\beta\gamma\delta\epsilon$
TAL015/020/024/028	TTGG	NG/NG/NN/NK	$\beta\gamma\delta\epsilon$
TAL015/020/024/029	TTGG	NG/NG/NN/NN	$\beta\gamma\delta\epsilon$
TAL015/020/024/030	TTGT	NG/NG/NN/NG	$\beta\gamma\delta\epsilon$
TAL015/020/025/026	TTTA	NG/NG/NG/NI	$\beta\gamma\delta\epsilon$
TAL015/020/025/027	TTTC	NG/NG/NG/HD	$\beta\gamma\delta\epsilon$
TAL015/020/025/028	TTTG	NG/NG/NG/NK	$\beta\gamma\delta\epsilon$
TAL015/020/025/029	TTTG	NG/NG/NG/NN	$\beta\gamma\delta\epsilon$
TAL015/020/025/030	TTTT	NG/NG/NG/NG	$\beta\gamma\delta\epsilon$
TAL011/016	AA	NI/NI	$\beta\gamma$
TAL011/017	AC	NI/HD	$\beta\gamma$
TAL011/018	AG	NI/NK	$\beta\gamma$
TAL011/019	AG	NI/NN	$\beta\gamma$
TAL011/020	AT	NI/NG	$\beta\gamma$
TAL012/016	CA	HD/NI	$\beta\gamma$
TAL012/017	CC	HD/HD	$\beta\gamma$
TAL012/018	CG	HD/NK	$\beta\gamma$
TAL012/019	CG	HD/NN	$\beta\gamma$
TAL012/020	CT	HD/NG	$\beta\gamma$
TAL013/016	GA	NK/NI	$\beta\gamma$

TAL013/017	GC	NK/HD	$\beta\gamma$
TAL013/018	GG	NK/NK	$\beta\gamma$
TAL013/019	GG	NK/NN	$\beta\gamma$
TAL013/020	GT	NK/NG	$\beta\gamma$
TAL014/016	GA	NN/NI	$\beta\gamma$
TAL014/017	GC	NN/HD	$\beta\gamma$
TAL014/018	GG	NN/NK	$\beta\gamma$
TAL014/019	GG	NN/NN	$\beta\gamma$
TAL014/020	GT	NN/NG	$\beta\gamma$
TAL015/016	TA	NG/NI	$\beta\gamma$
TAL015/017	TC	NG/HD	$\beta\gamma$
TAL015/018	TG	NG/NK	$\beta\gamma$
TAL015/019	TG	NG/NN	$\beta\gamma$
TAL015/020	TT	NG/NG	$\beta\gamma$
TAL011/016/021	AAA	NI/NI/NI	$\beta\gamma\delta$
TAL011/016/022	AAC	NI/NI/HD	$\beta\gamma\delta$
TAL011/016/023	AAG	NI/NI/NK	$\beta\gamma\delta$
TAL011/016/024	AAG	NI/NI/NN	$\beta\gamma\delta$
TAL011/016/025	AAT	NI/NI/NG	$\beta\gamma\delta$
TAL011/017/021	ACA	NI/HD/NI	$\beta\gamma\delta$
TAL011/017/022	ACC	NI/HD/HD	$\beta\gamma\delta$
TAL011/017/023	ACG	NI/HD/NK	$\beta\gamma\delta$
TAL011/017/024	ACG	NI/HD/NN	$\beta\gamma\delta$
TAL011/017/025	ACT	NI/HD/NG	$\beta\gamma\delta$
TAL011/018/021	AGA	NI/NK/NI	$\beta\gamma\delta$
TAL011/018/022	AGC	NI/NK/HD	$\beta\gamma\delta$
TAL011/018/023	AGG	NI/NK/NK	$\beta\gamma\delta$
TAL011/018/024	AGG	NI/NK/NN	$\beta\gamma\delta$
TAL011/018/025	AGT	NI/NK/NG	$\beta\gamma\delta$
TAL011/019/021	AGA	NI/NN/NI	$\beta\gamma\delta$
TAL011/019/022	AGC	NI/NN/HD	$\beta\gamma\delta$
TAL011/019/023	AGG	NI/NN/NK	$\beta\gamma\delta$
TAL011/019/024	AGG	NI/NN/NN	$\beta\gamma\delta$
TAL011/019/025	AGT	NI/NN/NG	$\beta\gamma\delta$
TAL011/020/021	ATA	NI/NG/NI	$\beta\gamma\delta$
TAL011/020/022	ATC	NI/NG/HD	$\beta\gamma\delta$
TAL011/020/023	ATG	NI/NG/NK	$\beta\gamma\delta$
TAL011/020/024	ATG	NI/NG/NN	$\beta\gamma\delta$
TAL011/020/025	ATT	NI/NG/NG	$\beta\gamma\delta$
TAL012/016/021	CAA	HD/NI/NI	$\beta\gamma\delta$
TAL012/016/022	CAC	HD/NI/HD	$\beta\gamma\delta$

TAL012/016/023	CAG	HD/NI/NK	$\beta\gamma\delta$
TAL012/016/024	CAG	HD/NI/NN	$\beta\gamma\delta$
TAL012/016/025	CAT	HD/NI/NG	$\beta\gamma\delta$
TAL012/017/021	CCA	HD/HD/NI	$\beta\gamma\delta$
TAL012/017/022	CCC	HD/HD/HD	$\beta\gamma\delta$
TAL012/017/023	CCG	HD/HD/NK	$\beta\gamma\delta$
TAL012/017/024	CCG	HD/HD/NN	$\beta\gamma\delta$
TAL012/017/025	CCT	HD/HD/NG	$\beta\gamma\delta$
TAL012/018/021	CGA	HD/NK/NI	$\beta\gamma\delta$
TAL012/018/022	CGC	HD/NK/HD	$\beta\gamma\delta$
TAL012/018/023	CGG	HD/NK/NK	$\beta\gamma\delta$
TAL012/018/024	CGG	HD/NK/NN	$\beta\gamma\delta$
TAL012/018/025	CGT	HD/NK/NG	$\beta\gamma\delta$
TAL012/019/021	CGA	HD/NN/NI	$\beta\gamma\delta$
TAL012/019/022	CGC	HD/NN/HD	$\beta\gamma\delta$
TAL012/019/023	CGG	HD/NN/NK	$\beta\gamma\delta$
TAL012/019/024	CGG	HD/NN/NN	$\beta\gamma\delta$
TAL012/019/025	CGT	HD/NN/NG	$\beta\gamma\delta$
TAL012/020/021	CTA	HD/NG/NI	$\beta\gamma\delta$
TAL012/020/022	CTC	HD/NG/HD	$\beta\gamma\delta$
TAL012/020/023	CTG	HD/NG/NK	$\beta\gamma\delta$
TAL012/020/024	CTG	HD/NG/NN	$\beta\gamma\delta$
TAL012/020/025	CTT	HD/NG/NG	$\beta\gamma\delta$
TAL013/016/021	GAA	NK/NI/NI	$\beta\gamma\delta$
TAL013/016/022	GAC	NK/NI/HD	$\beta\gamma\delta$
TAL013/016/023	GAG	NK/NI/NK	$\beta\gamma\delta$
TAL013/016/024	GAG	NK/NI/NN	$\beta\gamma\delta$
TAL013/016/025	GAT	NK/NI/NG	$\beta\gamma\delta$
TAL013/017/021	GCA	NK/HD/NI	$\beta\gamma\delta$
TAL013/017/022	GCC	NK/HD/HD	$\beta\gamma\delta$
TAL013/017/023	GCG	NK/HD/NK	$\beta\gamma\delta$
TAL013/017/024	GCG	NK/HD/NN	$\beta\gamma\delta$
TAL013/017/025	GCT	NK/HD/NG	$\beta\gamma\delta$
TAL013/018/021	GGA	NK/NK/NI	$\beta\gamma\delta$
TAL013/018/022	GGC	NK/NK/HD	$\beta\gamma\delta$
TAL013/018/023	GGG	NK/NK/NK	$\beta\gamma\delta$
TAL013/018/024	GGG	NK/NK/NN	$\beta\gamma\delta$
TAL013/018/025	GGT	NK/NK/NG	$\beta\gamma\delta$
TAL013/019/021	GGA	NK/NN/NI	$\beta\gamma\delta$
TAL013/019/022	GGC	NK/NN/HD	$\beta\gamma\delta$
TAL013/019/023	GGG	NK/NN/NK	$\beta\gamma\delta$

TAL013/019/024	GGG	NK/NN/NN	$\beta\gamma\delta$
TAL013/019/025	GGT	NK/NN/NG	$\beta\gamma\delta$
TAL013/020/021	GTA	NK/NG/NI	$\beta\gamma\delta$
TAL013/020/022	GTC	NK/NG/HD	$\beta\gamma\delta$
TAL013/020/023	GTG	NK/NG/NK	$\beta\gamma\delta$
TAL013/020/024	GTG	NK/NG/NN	$\beta\gamma\delta$
TAL013/020/025	GTT	NK/NG/NG	$\beta\gamma\delta$
TAL014/016/021	GAA	NN/NI/NI	$\beta\gamma\delta$
TAL014/016/022	GAC	NN/NI/HD	$\beta\gamma\delta$
TAL014/016/023	GAG	NN/NI/NK	$\beta\gamma\delta$
TAL014/016/024	GAG	NN/NI/NN	$\beta\gamma\delta$
TAL014/016/025	GAT	NN/NI/NG	$\beta\gamma\delta$
TAL014/017/021	GCA	NN/HD/NI	$\beta\gamma\delta$
TAL014/017/022	GCC	NN/HD/HD	$\beta\gamma\delta$
TAL014/017/023	GCG	NN/HD/NK	$\beta\gamma\delta$
TAL014/017/024	GCG	NN/HD/NN	$\beta\gamma\delta$
TAL014/017/025	GCT	NN/HD/NG	$\beta\gamma\delta$
TAL014/018/021	GGA	NN/NK/NI	$\beta\gamma\delta$
TAL014/018/022	GGC	NN/NK/HD	$\beta\gamma\delta$
TAL014/018/023	GGG	NN/NK/NK	$\beta\gamma\delta$
TAL014/018/024	GGG	NN/NK/NN	$\beta\gamma\delta$
TAL014/018/025	GGT	NN/NK/NG	$\beta\gamma\delta$
TAL014/019/021	GGA	NN/NN/NI	$\beta\gamma\delta$
TAL014/019/022	GGC	NN/NN/HD	$\beta\gamma\delta$
TAL014/019/023	GGG	NN/NN/NK	$\beta\gamma\delta$
TAL014/019/024	GGG	NN/NN/NN	$\beta\gamma\delta$
TAL014/019/025	GGT	NN/NN/NG	$\beta\gamma\delta$
TAL014/020/021	GTA	NN/NG/NI	$\beta\gamma\delta$
TAL014/020/022	GTC	NN/NG/HD	$\beta\gamma\delta$
TAL014/020/023	GTG	NN/NG/NK	$\beta\gamma\delta$
TAL014/020/024	GTG	NN/NG/NN	$\beta\gamma\delta$
TAL014/020/025	GTT	NN/NG/NG	$\beta\gamma\delta$
TAL015/016/021	TAA	NG/NI/NI	$\beta\gamma\delta$
TAL015/016/022	TAC	NG/NI/HD	$\beta\gamma\delta$
TAL015/016/023	TAG	NG/NI/NK	$\beta\gamma\delta$
TAL015/016/024	TAG	NG/NI/NN	$\beta\gamma\delta$
TAL015/016/025	TAT	NG/NI/NG	$\beta\gamma\delta$
TAL015/017/021	TCA	NG/HD/NI	$\beta\gamma\delta$
TAL015/017/022	TCC	NG/HD/HD	$\beta\gamma\delta$
TAL015/017/023	TCG	NG/HD/NK	$\beta\gamma\delta$
TAL015/017/024	TCG	NG/HD/NN	$\beta\gamma\delta$

TAL015/017/025	TCT	NG/HD/NG	$\beta\gamma\delta$
TAL015/018/021	TGA	NG/NK/NI	$\beta\gamma\delta$
TAL015/018/022	TGC	NG/NK/HD	$\beta\gamma\delta$
TAL015/018/023	TGG	NG/NK/NK	$\beta\gamma\delta$
TAL015/018/024	TGG	NG/NK/NN	$\beta\gamma\delta$
TAL015/018/025	TGT	NG/NK/NG	$\beta\gamma\delta$
TAL015/019/021	TGA	NG/NN/NI	$\beta\gamma\delta$
TAL015/019/022	TGC	NG/NN/HD	$\beta\gamma\delta$
TAL015/019/023	TGG	NG/NN/NK	$\beta\gamma\delta$
TAL015/019/024	TGG	NG/NN/NN	$\beta\gamma\delta$
TAL015/019/025	TGT	NG/NN/NG	$\beta\gamma\delta$
TAL015/020/021	TTA	NG/NG/NI	$\beta\gamma\delta$
TAL015/020/022	TTC	NG/NG/HD	$\beta\gamma\delta$
TAL015/020/023	TTG	NG/NG/NK	$\beta\gamma\delta$
TAL015/020/024	TTG	NG/NG/NN	$\beta\gamma\delta$
TAL015/020/025	TTT	NG/NG/NG	$\beta\gamma\delta$
TAL011/031	AA	NI/NI	$\beta\gamma'$
TAL011/032	AC	NI/HD	$\beta\gamma'$
TAL011/033	AG	NI/NK	$\beta\gamma'$
TAL011/034	AG	NI/NN	$\beta\gamma'$
TAL011/035	AT	NI/NG	$\beta\gamma'$
TAL012/031	CA	HD/NI	$\beta\gamma'$
TAL012/032	CC	HD/HD	$\beta\gamma'$
TAL012/033	CG	HD/NK	$\beta\gamma'$
TAL012/034	CG	HD/NN	$\beta\gamma'$
TAL012/035	CT	HD/NG	$\beta\gamma'$
TAL013/031	GA	NK/NI	$\beta\gamma'$
TAL013/032	GC	NK/HD	$\beta\gamma'$
TAL013/033	GG	NK/NK	$\beta\gamma'$
TAL013/034	GG	NK/NN	$\beta\gamma'$
TAL013/035	GT	NK/NG	$\beta\gamma'$
TAL014/031	GA	NN/NI	$\beta\gamma'$
TAL014/032	GC	NN/HD	$\beta\gamma'$
TAL014/033	GG	NN/NK	$\beta\gamma'$
TAL014/034	GG	NN/NN	$\beta\gamma'$
TAL014/035	GT	NN/NG	$\beta\gamma'$
TAL015/031	TA	NG/NI	$\beta\gamma'$
TAL015/032	TC	NG/HD	$\beta\gamma'$
TAL015/033	TG	NG/NK	$\beta\gamma'$
TAL015/034	TG	NG/NN	$\beta\gamma'$
TAL015/035	TT	NG/NG	$\beta\gamma'$

TAL021/036	AA	NI/NI	$\delta\epsilon'$
TAL021/037	AC	NI/HD	$\delta\epsilon'$
TAL021/038	AG	NI/NK	$\delta\epsilon'$
TAL021/039	AG	NI/NN	$\delta\epsilon'$
TAL021/040	AT	NI/NG	$\delta\epsilon'$
TAL022/036	CA	HD/NI	$\delta\epsilon'$
TAL022/037	CC	HD/HD	$\delta\epsilon'$
TAL022/038	CG	HD/NK	$\delta\epsilon'$
TAL022/039	CG	HD/NN	$\delta\epsilon'$
TAL022/040	CT	HD/NG	$\delta\epsilon'$
TAL023/036	GA	NK/NI	$\delta\epsilon'$
TAL023/037	GC	NK/HD	$\delta\epsilon'$
TAL023/038	GG	NK/NK	$\delta\epsilon'$
TAL023/039	GG	NK/NN	$\delta\epsilon'$
TAL023/040	GT	NK/NG	$\delta\epsilon'$
TAL024/036	GA	NN/NI	$\delta\epsilon'$
TAL024/037	GC	NN/HD	$\delta\epsilon'$
TAL024/038	GG	NN/NK	$\delta\epsilon'$
TAL024/039	GG	NN/NN	$\delta\epsilon'$
TAL024/040	GT	NN/NG	$\delta\epsilon'$
TAL025/036	TA	NG/NI	$\delta\epsilon'$
TAL025/037	TC	NG/HD	$\delta\epsilon'$
TAL025/038	TG	NG/NK	$\delta\epsilon'$
TAL025/039	TG	NG/NN	$\delta\epsilon'$
TAL025/040	TT	NG/NG	$\delta\epsilon'$
TAL011	A	NI	β
TAL012	C	HD	β
TAL013	G	NK	β
TAL014	G	NN	β
TAL015	T	NG	β

To prepare DNA fragments encoding α units for use in assembly, 20 rounds of PCR were performed with each α unit plasmid as a template using primers oJS2581 (5'-Biotin-TCTAGAGAAGACAAGAACCTGACC-3' (SEQ ID NO:237)) and oJS2582 (5'-GGATCCGGTCTCTTAAGGCCGTGG-3' (SEQ ID NO:238)). The resulting PCR products were biotinylated on the 5' end. Each α PCR product was then digested with 40 units of BsaI-HF restriction enzyme to generate 4bp overhangs, purified using the

QIAquick PCR purification kit (QIAGEN) according to manufacturer's instructions except that the final product was eluted in 50 μ l of 0.1X EB.

To prepare DNA fragments encoding β , $\beta\gamma\delta\epsilon$, $\beta\gamma\delta$, $\beta\gamma$, $\beta\gamma^*$, and $\delta\epsilon^*$ repeats, 10 μ g of each of these plasmids was digested with 50 units of BbsI restriction enzyme in NEBuffer 2 for 2 hours at 37 °C followed by serial restriction digests performed in NEBuffer 4 at 37 °C using 100 units each of XbaI, BamHI-HF, and Sall-HF enzymes that were added at 5 minute intervals. The latter set of restriction digestions were designed to cleave the plasmid backbone to ensure that this larger DNA fragment does not interfere with subsequent ligations performed during the assembly process. These restriction digest reactions were then purified using the QIAquick PCR purification kit (QIAGEN) according to manufacturer's instructions except that the final product was eluted in 180 μ l of 0.1X EB.

All assembly steps were performed using a Sciclone G3 liquid handling workstation (Caliper) in 96-well plates and using a SPRIplate 96-ring magnet (Beckman Coulter Genomics) and a DynaMag-96 Side magnet (Life Technologies). In the first assembly step, a biotinylated α unit fragment was ligated to the first $\beta\gamma\delta\epsilon$ fragment and then the resulting $\alpha\beta\gamma\delta\epsilon$ fragments are bound to Dynabeads MyOne C1 streptavidin-coated magnetic beads (Life Technologies) in 2X B&W Buffer (Life Technologies). Beads were then drawn to the side of the well by placing the plate on the magnet and then washed with 100 μ l B&W buffer with 0.005% Tween 20 (Sigma) and again with 100 μ l 0.1 mg/ml bovine serum albumin (BSA) (New England Biolabs). Additional $\beta\gamma\delta\epsilon$ fragments were ligated by removing the plate from the magnet, resuspending the beads in solution in each well, digesting the bead bound fragment with BsaI-HF restriction enzyme, placing the plate on the magnet, washing with 100 μ l B&W/Tween20 followed by 100 μ l of 0.1 mg/ml BSA, and then ligating the next fragment. This process was repeated multiple times with additional $\beta\gamma\delta\epsilon$ units to extend the bead-bound fragment. The last fragment to be ligated was always a β , $\beta\gamma^*$, $\beta\gamma\delta$, or $\delta\epsilon^*$ unit to enable cloning of the full-length fragment into expression vectors (note that fragments that end with a $\delta\epsilon^*$ unit are always preceded by ligation of a $\beta\gamma$ unit).

The final full-length bead-bound fragment was digested with 40 units of BsaI-HF restriction enzyme followed by 25 units of BbsI restriction enzyme (New England Biolabs). Digestion with BbsI released the fragment from the beads and generated a unique 5' overhang for cloning of the fragment. Digestion with BsaI-HF resulted in
5 creation of a unique 3' overhang for cloning.

DNA fragments encoding the assembled TALE repeat arrays were subcloned into one of four TALEN expression vectors. Each of these vectors included a CMV promoter, a translational start codon optimized for mammalian cell expression, a triple FLAG epitope tag, a nuclear localization signal, amino acids 153 to 288 from the TALE 13
10 protein (Miller et al., 2011, Nat. Biotechnol., 29:143-148), two unique and closely positioned Type IIS BsmBI restriction sites, a 0.5 TALE repeat domain encoding one of four possible RVDs (NI, HD, NN, or NG for recognition of an A, C, G, or T nucleotide, respectively), amino acids 715 to 777 from the TALE 13 protein, and the wild-type FokI cleavage domain. All DNA fragments possessed overhangs that enable directional
15 cloning into any of the four TALEN expression vectors that has been digested with BsmBI.

To prepare a TALEN expression vector for subcloning, 5 µg of plasmid DNA were digested with 50 units of BsmBI restriction enzyme (New England Biolabs) in NEBuffer 3 for 8 hours at 55 degrees C. Digested DNA was purified using 90 µl of
20 Ampure XP beads (Agencourt) according to manufacturer's instructions and diluted to a final concentration of 5ng/µl in 1 mM TrisHCl. The assembled TALE repeat arrays were ligated into TALEN expression vectors using 400 U of T4 DNA Ligase (New England Biolabs). Ligation products were transformed into chemically competent XL-1 Blue cells. Six colonies were picked for each ligation and plasmid DNA isolated by an
25 alkaline lysis miniprep procedure. Simultaneously, the same six colonies were screened by PCR using primers oSQT34 (5'-GACGGTGGCTGTCAAATACCAAGATATG-3' (SEQ ID NO:239)) and oSQT35 (5'-TCTCCTCCAGTTCACCTTTGACTAGTTGGG - 3' (SEQ ID NO:240)). PCR products were analyzed on a QIAxcel capillary electrophoresis system (Qiagen). Miniprep DNA from clones that contained correctly
30 sized PCR products were sent for DNA sequence confirmation with primers oSQT1 (5'-

AGTAACAGCGGTAGAGGCAG-3' (SEQ ID NO:241)), oSQT3
(5'-ATTGGGCTACGATGGACTCC-3' (SEQ ID NO:242)), and oJS2980
(5'-TTAATTCAATATATTCATGAGGCAC-3' (SEQ ID NO:243)).

5 Because the final fragment ligated can encode one, two, or three TALE repeats,
the methods disclosed herein can be used to assemble arrays consisting of any desired
number of TALE repeats. Assembled DNA fragments encoding the final full-length
TALE repeat array are released from the beads by restriction enzyme digestion and can
be directly cloned into a desired expression vector of choice.

10 The methods can be efficiently practiced in 96-well format using a robotic liquid
handling workstation. With automation, DNA fragments encoding 96 different TALE
repeat arrays of variable lengths can be assembled in less than one day. Medium-
throughput assembly of fragments can be performed in one to two days using multi-
channel pipets and 96-well plates. Fragments assembled using either approach can then
be cloned into expression vectors (e.g., for expression as a TALEN) to generate sequence-
15 verified plasmids in less than one week. Using the automated assembly approach,
sequence-verified TALE repeat array expression plasmids can be made quickly and
inexpensively.

20 Example 6. Large-scale Testing of Assembled TALENs Using a Human Cell-based
Reporter Assay

To perform a large-scale test of the robustness of TALENs for genome editing in
human cells, the method described in Example 5 was used to construct a series of
plasmids encoding 48 TALEN pairs targeted to different sites scattered throughout the
EGFP reporter gene. Monomers in each of the TALEN pairs contained the same number
25 of repeats (ranging from 8.5 to 19.5 in number), and these pairs were targeted to sites
possessing a fixed length "spacer" sequence (16 bps) between the "half-sites" bound by
each TALEN monomer (Table 6).

Table 6. EGFP reporter gene sequences targeted by 48 pairs of TALENs

TALEN pair #	Position within EGFP of the first nucleotide in the binding site	Target site (half-sites in CAPS, spacer in lowercase)	SEQ ID NO:	# of repeat domains in Left TALEN	# of repeat domains in Right TALEN
1	-8	TCGCCACCATggtgagcaaggg cgagGAGCTGTTCA	93	8.5	8.5
2	35	TGGTGGCCATcctggtcgagct ggacGGCGACGTAA	94	8.5	8.5
3	143	TCTGCACCACgggcaagctgcc cgtgCCCTGGCCCA	95	8.5	8.5
4	425	TGGAGTACAActacaacagcca caacGTCTATATCA	96	8.5	8.5
5	82	TTCAGCGTGTcggcgagggcg agggcGATGCCACCTA	97	9.5	9.5
6	111	TGCCACCTACGgcaagctgacc ctgaaGTCATCTGCA	98	9.5	9.5
7	172	TGGCCCACCCTcgtgaccaccc tgaccTACGGCGTGCA	99	9.5	9.5
8	496	TTCAAGATCCGccacaacatcg aggacGGCAGCGTGCA	100	9.5	9.5
9	-23	TAGAGGATCCACcggctcgccac catggtGAGCAAGGGCGA	101	10.5	10.5
10	91	TCCGGCGAGGGCgagggcgatg ccacctACGGCAAGCTGA	102	10.5	10.5
11	194	TGACCTACGGCGtgagtgtt cagccgTACCCGACCA	103	10.5	10.5
12	503	TCCGCCACAACatcgaggacgg cagcgtGCAGCTCGCCGA	104	10.5	10.5
13	44	TCCTGGTCGAGCTggacggcga cgtaaacGGCCACAAGTTCA	105	11.5	11.5
14	215	TCAGCCGTACCCcgaccacat gaagcagCACGACTTCTTCA	106	11.5	11.5
15	251	TCTTCAAGTCCGCcatgccga aggctacTCCAGGAGCGCA	107	11.5	11.5
16	392	TCAAGGAGGACGGcaacatcct ggggcacAAGCTGGAGTACA	108	11.5	11.5
17	485	TCAAGGTGAACttcaagatccg ccacaacATCGAGGACGGCA	109	11.5	11.5
18	-16	TCCACCGGTCGCCAccatggtg agcaagggCGAGGAGCTGTTCA	110	12.5	12.5
19	82	TTCAGCGTGTCCGGcgagggcg agggcgatGCCACCTACGGCAA	111	12.5	12.5
20	214	TTCAGCCGTACCCcgaccaca tgaagcagCACGACTTCTTCAA	112	12.5	12.5
21	436	TACAACAGCCACAacgtctata tcatggccGACAAGCAGAAGAA	113	12.5	12.5

22	35	TGGTGGCCATCCTGGtcgagct ggacggcgaCGTAAACGGCCAC AA	114	13.5	13.5
23	266	TGCCCCAAGGCTACGtccagga gcgcaccatCTTCTTCAAGGAC GA	115	13.5	13.5
24	362	TGAACCGCATCGAGCtgaagg catcgacttCAAGGAGGACGGC AA	116	13.5	13.5
25	497	TCAAGATCCGCCACAacatcga ggacggcagCGTGCAGCTCGCC GA	117	13.5	13.5
26	23	TGTTACCGGGGTGGTgccc cctggtcgagCTGGACGGCGAC GTAA	118	14.5	14.5
27	38	TGCCCCATCCTGGTCGagctgga cggcgacgtaAACGGCCACAAG TTCA	119	14.5	14.5
28	89	TGTCCGGCGAGGGCGAgggcga tgccacctacGGCAAGCTGACC CTGA	120	14.5	14.5
29	140	TCATCTGCACCACCGGcaagct gcccgtgcccTGGCCCACCCTC GTGA	121	14.5	14.5
30	452	TCTATATCATGGCCGAcaagca gaagaacggcATCAAGGTGAAC TTCA	122	14.5	14.5
31	199	TACGGCGTGCACTGCTTcagcc gctaccccgacCACATGAAGCA GCACGA	123	15.5	15.5
32	223	TACCCCGACCACATGAAGcagc acgacttcttcAAGTCCGCCAT GCCCGA	124	15.5	15.5
33	259	TCCGCCATGCCCGAAGGctacg tccaggagcgcACCATCTTCTT CAAGGA	125	15.5	15.5
34	391	TTCAAGGAGGACGGCAacatcc tggggcacaagCTGGAGTACAA CTACAA	126	15.5	15.5
35	430	TACAACTACAACAGCCaacaag tctatatcatgGCCGACAAGCA GAAGAA	127	15.5	15.5
36	26	TCACCGGGGTGGTGCCCAtcct ggtcgagctggaCGGCGACGTA AACGGCCA	128	16.5	16.5
37	68	TAAACGGCCACAAGTTCAgcgt gtccggcgaggCGAGGGCGAT GCCACCTA	129	16.5	16.5
38	206	TGCAGTGCTTCAGCCGCTacc cgaccacatgaaGCAGCACGAC TTCTTCAA	130	16.5	16.5
39	83	TCAGCGTGTCGGCGAGGGcga gggcgatgccaccTACGGCAAG CTGACCCTGA	131	17.5	17.5

40	134	TGAAGTTCATCTGCACCACcgg caagctgcccgtgCCCTGGCCC ACCCCTCGTGA	132	17.5	17.5
41	182	TCGTGACCACCCTGACCTAcgg cgtgcagtgttcAGCCGCTAC CCCGACCACA	133	17.5	17.5
42	458	TCATGGCCGACAAGCAGAAgaa cggcatcaaggtgAACTTCAAG ATCCGCCACA	134	17.5	17.5
43	25	TTCACCGGGTGGTGCCCATcc tggtcgagctggacGGCGACGT AAACGCCACAA	135	18.5	18.5
44	145	TGCACCACCGGCAAGCTGCCcg tgccctggcccaccCTCGTGAC CACCTGACCTA	136	18.5	18.5
45	253	TTCAAGTCCGCCATGCCCGAag gctacgtccaggagCGCACCAT CTTCTTCAAGGA	137	18.5	18.5
46	454	TATATCATGGCCGACAAGCAga agaacggcatcaagGTGAACCTT CAAGATCCGCCA	138	18.5	18.5
47	139	TTCATCTGCACCACCGGCAAGc tgcccgtgccctggcCCACCCT CGTGACCACCCTGA	139	19.5	19.5
48	338	TGAAGTTCGAGGGCGACACCct ggtgaaccgcatcgaGCTGAAG GGCATCGACTTCAA	140	19.5	19.5

Each of the 48 TALEN pairs was tested in human cells for its ability to disrupt the coding sequence of a chromosomally integrated EGFP reporter gene. In this assay, NHEJ-mediated repair of TALEN-induced breaks within the EGFP coding sequence led to loss of EGFP expression, which was quantitatively assessed using flow cytometry 2 and 5 days following transfection. (To ensure that activities of each active TALEN pair could be detected, we only targeted sites located at or upstream of nucleotide position 503 in the gene, a position we had previously shown would disrupt EGFP function when mutated with a zinc finger nuclease (ZFN) (Maeder et al., 2008, Mol. Cell 31:294-301).) Strikingly, all 48 TALEN pairs showed significant EGFP gene-disruption activities in this assay (FIG. 19A). The net percentage of EGFP-disrupted cells induced by TALENs on day 2 post-transfection ranged from 9.4% to 68.0%, levels comparable to the percentage disruption observed with four EGFP-targeted ZFN pairs originally made by the Oligomerized Pool Engineering (OPEN) method (Fig. 19A). These results demonstrate that TALENs containing as few as 8.5 TALE repeats possess significant nuclease

activities and provide a large-scale demonstration of the robustness of TALENs in human cells.

Interestingly, re-quantification of the percentage of EGFP-negative cells at day 5 post-transfection revealed that cells expressing shorter-length TALENs (such as those composed of 8.5 to 10.5 repeats) showed significant reductions in the percentage of EGFP-disrupted cells whereas those expressing longer TALENs did not (FIGs. 19A-B and 20A). One potential explanation for this effect is cellular toxicity associated with expression of shorter-length TALENs. Consistent with this hypothesis, in cells transfected with plasmids encoding shorter-length TALENs, greater reductions in the percentage of tdTomato-positive cells were observed from day 2 to day 5 post-transfection (FIG. 20D) (a tdTomato-encoding plasmid was co-transfected together with the TALEN expression plasmids on day 0). Taken together, our results suggest that although shorter-length TALENs are as active as longer-length TALENs, the former can cause greater cytotoxicity in human cells.

Our EGFP experiments also provided an opportunity to assess four of five computationally-derived design guidelines (Cermak et al., 2011, *Nucleic Acids Res.*, 39:e82). The guidelines proposed by Cermak are as follows:

1. The nucleotide just 5' to the first nucleotide of the half-site should be a thymine.
2. The first nucleotide of the half-site should not be a thymine.
3. The second nucleotide of the half-site should not be an adenosine.
4. The 3' most nucleotide in the target half-site should be a thymine.
5. The composition of each nucleotide within the target half-site should not vary from the observed percentage composition of naturally occurring binding sites by more than 2 standard deviations. The percentage composition of all naturally occurring TALE binding sites is: A = 31±16%, C = 37±13%, G = 9±8%, T = 22±10%. Hence, the nucleotide composition of potential TALE binding sites should be: A = 0% to 63%, C = 11% to 63%, G = 0% to 25% and T = 2% to 42%.

These guidelines have been implemented in the TALE-NT webserver (boglabx.plp.iastate.edu/TALENT/TALENT/) to assist users in identifying potential

TALEN target sites. All 48 of the sequences we targeted in EGFP did not meet one or more of these guidelines (however, note that all of our sites did meet the requirement for a 5' T). The ~100% success rate observed for these 48 sites demonstrates that TALENs can be readily obtained for target sequences that do not follow these guidelines. In addition, for each of the four design guidelines, we did not find any statistically significant correlation between guideline violation and the level of TALEN-induced mutagenesis on either day 2 or day 5 post-transfection. We also failed to find a significant correlation between the total number of guideline violations and the level of mutagenic TALEN activity. Thus, our results show that failure to meet four of the five previously described design guidelines when identifying potential TALEN target sites does not appear to adversely affect success rates or nuclease efficiencies.

Example 7. High-throughput Alteration of Endogenous Human Genes Using Assembled TALENs

Having established the robustness of the TALEN platform with a chromosomally integrated reporter gene, it was next determined whether this high success rate would also be observed with endogenous genes in human cells. To test this, the assembly method described in Example 5 was used to engineer TALEN pairs targeted to 96 different human genes: 78 genes implicated in human cancer (Vogelstein and Kinzler, 2004, Nat. Med., 10:789-799) and 18 genes involved in epigenetic regulation of gene expression (Table 7). For each gene, a TALEN pair was designed to cleave near the amino-terminal end of the protein coding sequence, although in a small number of cases the presence of repetitive sequences led us to target alternate sites in neighboring downstream exons or introns (Table 7). Guided by the results with the EGFP TALENs, TALENs composed of 14.5, 15.5, or 16.5 repeats were constructed that cleaved sites with 16, 17, 18, 19 or 21 bp spacer sequences. All of the target sites had a T at the 5' end of each half-site.

Table 7. Endogenous human gene sequences targeted by 96 pairs of TALENs

Target gene name	% NHEJ	Target site (half-sites in CAPS, spacer in lowercase, ATG underlined)	SEQ ID NO:	Length of LEFT half site (include 5' T)	Length of spacer	Length of RIGHT half site (include 5' T)	Gene Type
ABL1	22.5 ± 7.1	TACCTATTATTACT TTATggggcagcagcctgg aaAAGTACTTGGGG ACCAA	141.	16.5	17	15.5	Cancer
AKT2	14.1 ± 7.3	TGTGTCTTGGGATG AGTGggtcagtggttggtg CTCACAGGATGGCT GGCA	142.	16.5	16	16.5	Cancer
ALK	12.7 ± 2.9	TCCTGTGGCTCCTG CCGCtgetgettccagge AGCTGTGGGCTCCG GGA	143.	16.5	16	15.5	Cancer
APC	48.8 ± 9.8	TATGTACGCCTCCC TGGGctcgggtccggctgcc CCTTTGCCCGCTTC TGTA	144.	16.5	16	16.5	Cancer
ATM	35.5 ± 15.6	TGAATTGGGATGCT GTTTttaggtattctattcaaa TTTATTTTACTGTCT TTA	145.	16.5	18	16.5	Cancer
AXIN2	2.5 ± 0.6	TCCCTCACCATGAG TAGCgctatgttggtgacttG CCTCCCGGACCCCA GCA	146.	16.5	16	16.5	Cancer
BAX	14.7 ± 11.6	TGTGCGATCTCCAA GCACtgaggggcagaaact cCCGGATCGGGCGC TGCCA	147.	16.5	16	16.5	Cancer
BCL6	14.9 ± 5.9	TTTTCAAGTGAAGA CAAAatggcctcgccggct gACAGCTGTATCCA GTTCA	148.	16.5	16	16.5	Cancer
BMPR1 A	50.4 ± 16.4	TACAATTGAACAAT GCCTeagctatacattacat CAGATTATTGGGAG CCTA	149.	16.5	17	16.5	Cancer
BRCA1	44.5 ± 15.5	TCCGAAGCTGACAG ATGGgtattcttgacgggg GGTAGGGGCGGAA CCTGA	150.	16.5	16	16.5	Cancer
BRCA2	41.6 ± 10.5	TTAGACTTAGGTAA GTAAtgcataatggtagact GGGGAGAACTACA AACTA	151.	16.5	16	16.5	Cancer

CBX3	35.2 ± 22.6	TCTGCAATAAAAAA TGGCctccaacaaactaca TTGGTAAGTTAATG AAAA	152.	16.5	16	16.5	Epigenetic
CBX8	13.5 ± 3.4	TGGAGCTTTCAGCG GTGGgggagcgggtgttcg cgGCCGAAGCCCTC CTGAA	153.	16.5	17	15.5	Epigenetic
CCND1	40.5 ± 2.2	TGGAACACCAGCTC CTGTgctgcgaagtggaaac catCCGCCGCGCGTA CCCCGA	154.	16.5	19	16.5	Cancer
CDC73	36.3 ± 7.7	TGCTTAGCGTCCTG CGACagtacaacatccagaa GAAGGAGATTGTG GTGAA	155.	16.5	16	16.5	Cancer
CDH1	none	TGCTGCAGGTACCC CGGAteccctgacttgcgag GGACGCATTTCGGGC CGCA	156.	16.5	16	16.5	Cancer
CDK4	21.5 ± 17.4	TCCCTTGATCTGAG AAtggctacctctcgataTG AGCCAGTGGCTGA AA	157.	14.5	16	15.5	Cancer
CHD4	9.6 ± 0.1	TGGCGTCGGGCCTG GGCtccccgtccccctgctc GGCGGGCAGTGAG GAGGA	158.	15.5	17	16.5	Epigenetic
CHD7	11.4 ± 2.7	TGTGTTGGAAGAAG ATGGcagatccaggaatgat GAGTCTTTTTGGCG AGGA	159.	16.5	16	16.5	Epigenetic
CTNNB1	26.0 ± 8.1	TCCAGCGTGGACAA TGGctactcaagggttgTC ATTAAATCTTTAGT TA	160.	15.5	16	16.5	Cancer
CYLD	24.7 ± 2.3	TAATATCACAATGA GTTcaggcttatggagccaa gaAAAAGTCACTTC ACCCTA	161.	16.5	18	16.5	Cancer
DDB2	15.8 ± 7.2	TCACACGGAGGAC GCGatgctccaagaagaaac GCCCAGAAACCCA GAAGA	162.	14.5	16	16.5	Cancer
ERCC2	55.8 ± 12.7	TCCGGCCGGGCCCA TGAagtgagaagggggtg GGGGTCGCGCTCGC TA	163.	15.5	16	14.5	Cancer
ERCC5	none	TCCGGGATCGCCAT GGGAactcaatagaaaatcc teaTCTTCTCACTTTG TTCA	164.	16.5	19	16.5	Cancer

EWSR1	14.3 ± 8.2	TGGCGTCCACGGGT GAGTatggtggaactgcggt cGCGCCGGCGGTAG CCGGA	165.	16.5	17	16.5	Cancer
EXT1	9.5 ± 3.0	TGACCCAGGCAGG ACACA tgcaggccaaaaa cgcTATTTTCATCCTG CTCTCA	166.	16.5	17	16.5	Cancer
EXT2	none	TTCCTCCCAGGGGG ATGTcctgcgcctcagggtc CGGTGGTGGCCTGC GGCA	167.	16.5	16	16.5	Cancer
EZH2	41.3 ± 2.6	TGCTTTTAGAATAA TCATgggcccagactgggaa gAAATCTGAGAAGG GACCA	168.	16.5	16	16.5	Epigenetic
FANCA	9.7 ± 5.0	TAGGCGCCAAGGC CATGTccgactcgtgggtc ccGAACTCCGCCTC GGGCA	169.	16.5	16	16.5	Cancer
FANCC	23.7 ± 17.8	TGAAGGGACATCA CCTTTcgtttttccaagatg GCTCAAGATTCAAGT AGA	170.	16.5	17	15.5	Cancer
FANCE	none	TGCCCCGGCATGGC GACAccggagcgggggtc ccTGGGGCTGAGGG CGTGA	171.	16.5	17	16.5	Cancer
FANCF	46.0 ± 7.7	TTCGCGCACCTCAT GGaatcccttctgcagcaCC TGGATCGCTTTTCC GA	172.	14.5	16	16.5	Cancer
FANCG	26.9 ± 16.2	TCGGCCACCATGTC CCgccagaccactctgtGG GCTCCAGCTGCCTG GA	173.	14.5	16	16.5	Cancer
FES	12.6 ± 10.6	TCCCCAGAACAGCA CTATgggcttctctccgagc tGTGCAGCCCCCAG GGCCA	174.	16.5	18	16.5	Cancer
FGFR1	17.4 ± 6.2	TCTGCTCCCCACCG AGGAacctctgcagcgca TGAATCCCAGGAGC CTA	175.	16.5	16	15.5	Cancer
FH	20.9 ± 11.8	TGTACCGAGCACTT CGGCtctcgcgcgctcgcg tCCCCTCGTGCGGG CTCCA	176.	16.5	17	16.5	Cancer
FLCN	11.1 ± 4.4	TCTCCAAGGCACCA TGAAtgccatcgtggtctct gCCACTTCTGCGAG CTCCA	177.	16.5	18	16.5	Cancer

FLT3	none	TCCGGAGGCCATGC CGGCgttgccgcgcgacgg cggccaGCTGCCGCTG CTCGGTA	178.	16.5	21	15.5	Cancer
FLT4	9.9 ± 5.0	TGCAGCGGGGCGC CGCGCtgtgcctgcgactgt ggctCTGCCTGGGAC TCCTGGA	179.	16.5	19	16.5	Cancer
FOXO1	8.5 ± 1.1	TCACCATGGCCGAG GCGcctcaggtggaggagaT CGACCCGGACTTCG A	180.	15.5	16	14.5	Cancer
FOXO3	7.3 ± 2.3	TCTCCGCTCGAAGT GGA Gctggacccggagttc gagCCCCAGAGCCGT CCGCGA	181.	16.5	18	16.5	Cancer
GLI1	21.5 ± 12.4	TCCTCTGAGACGCC ATGTtcaactcgatgacccc ACCACCAATCAGTA GCTA	182.	16.5	16	16.5	Cancer
HDAC1	10.8 ± 3.0	TGGCGCAGACGCA GGGCacccggaggaaagtc tgTTACTACTACGAC GGTGA	183.	15.5	17	16.5	Epigen etic
HDAC2	4.2 ± 0.9	TGCGCTCACCTCCC TGCGgcctcctgaggtggtt gGTGGCCCCCTCCT CGCGA	184.	16.5	18	16.5	Epigen etic
HDAC6	21.4 ± 2.1	TCCTCAACTATGAC CTCAacggccaggattcca CCACAACCAGGCA GCGAA	185.	16.5	16	16.5	Epigen etic
HMGA 2	3.0 ± 1.5	TGAGCGCACGCGGT GAGGgcgcggggcagccg tcCACTTCAGCCCAG GGACA	186.	16.5	16	16.5	Cancer
HOXA1 3	7.6 ± 3.1	TCCGTGCTCCTCCA CCCCcgtggatcgagccca cCGTCATGTTTCTCT ACGA	187.	16.5	17	16.5	Cancer
HOXA9	6.4 ± 2.7	TGGGCACGGTGATG GCcaccactggggccctgG GCAACTACTACGTG GA	188.	14.5	16	15.5	Cancer
HOXC1 3	10.5 ± 0.3	TCCAGCAGATCATG TCATgacgacttcgctgctcc tGCATCCACGCTGG CCGGA	189.	16.5	18	16.5	Cancer
HOXD1 1	none	TTGACGAGTGCGGC CAGagcgcagccagcatgta CCTGCCGGGCTGCG CCTA	190.	15.5	17	16.5	Cancer

HOXD1 3	none	TGCGGGCAGACGG CGGGGgcgccggtggcgc cccGcCTCTTCCTCC TCCTCA	191.	16.5	17	16.5	Cancer
JAK2	44.9 ± 16.9	TCTGAAAAAGACTC TGCAtggaatggcctgct TACGATGACAGAA ATGGA	192.	16.5	16	16.5	Cancer
KIT	none	TACCGCGATGAGA GGCGCtcgcggcgctgg gattttCTCTGCGTTCT GCTCCTA	193.	16.5	19	16.5	Cancer
KRAS	9.4 ± 0.9	TGAAAATGACTGA ATATAaacttggtagt gaGCTGGTGGCGTA GGCAA	194.	16.5	17	15.5	Cancer
MAP2K 4	11.9 ± 7.1	TAGGGTCCCCGGCG CCAGgccacccggccgta gCAGCATGCAGGGT AAGGA	195.	16.5	16	16.5	Cancer
MDM2	33.0 ± 20.2	TCCAAGCGCGAAA ACCCCggatggtgaggag caggTACTGGCCCCG CAGCGA	196.	16.5	17	15.5	Cancer
MET	40.4 ± 10.7	TTATTATTACATGG CTTTgccttactgagcttcA TCTTGTCTCTGGT CCA	197.	16.5	16	16.5	Cancer
MLH1	44.9 ± 6.3	TCTGGCGCCAAAAT GTCGttcttggcagggtta TTCGGCGGCTGGAC GAGA	198.	16.5	16	16.5	Cancer
MSH2	27.5 ± 10.4	TGAGGAGGTTTCGA CATGgcggtgcagcgaag gAGACGCTGCAGTT GGAGA	199.	16.5	16	16.5	Cancer
MUTY H	24.9 ± 8.4	TCACTGTCGGCGGC CATGacacgcctcgtctccc gcCTGAGTCGTCTGT GGGTA	200.	16.5	18	16.5	Cancer
MYC	13.4 ± 4.0	TGCTTAGACGCTGG ATTTtttccggtagtgaaA ACCAGGTAAGCAC CGAA	201.	16.5	16	16.5	Cancer
MYCL1	17.3 ± 0.6	TCCCGCAGGGAGC GGACAtggactacgactcg taCCAGCACTATTTC TACGA	202.	16.5	16	16.5	Cancer
MYCN	16.3 ± 11.6	TGCCGAGCTGCTCC ACgtccaccatgccgggcA TGATCTGCAAGAAC CCA	203.	14.5	16	16.5	Cancer

NBN	46.3 ± 15.5	TGAGGAGCCGGAC CGAtgtgaaactgtgccC GCCGCGGGCCCGG CA	204.	14.5	16	14.5	Cancer
NCOR1	29.6 ± 13.1	TCTTTACTGATAAT GTCAagttcatgttaacctcC CAACCAAGGAGCA TTCA	205.	16.5	16	16.5	Epigen etic
NCOR2	3.3 ± 0.6	TGGAGGGCCACTG AGCcccgctaccgcccac CAGCCTTTCTACC CA	206.	14.5	16	14.5	Epigen etic
NTRK1	none	TCGGCGCATGAAG GAGGTactctcatittcgtt CTCTCTCTCTGTGC CCCA	207.	16.5	16	16.5	Cancer
PDGFR A	16.0 ± 4.3	TTGCGCTCGGGGCG GCCAtgtggcgccgagagg ICGAGCGCCTAGTG TCGGA	208.	16.5	16	16.5	Cancer
PDGFR B	16.0 ± 3.2	TCTGCAGGACACCA TGCGgcttcgggtgcgatg CCAGCTCTGGCCCT CAAA	209.	16.5	16	16.5	Cancer
PHF8	22.2 ± 6.1	TGAGTACTCCGCCT CTACcccggtgaagcccg cCCCCGCCGCCACC TATTA	210.	16.5	16	16.5	Epigen etic
PMS2	26.9 ± 9.5	TCGGGTGTTGCATC CATGgagcgagctgagagc tcgAGGTGAGCGGG GCTCGCA	211.	16.5	18	16.5	Cancer
PTCH1	27.5 ± 15.9	TGGAAGTCTTAAT AGaaacaggttgaattGT GAGTCCGCGCTGCA	212.	14.5	16	14.5	Cancer
PTEN	31.5 ± 11.7	TCCCAGACATGACA GCCatcatcaaagagatcgT TAGCAGAAACAAA AGGA	213.	15.5	16	16.5	Cancer
RARA	13.4 ± 6.1	TGGCATGGCCAGCA ACAGcagctctgcccgcac acCTGGGGGCGGGC ACCTCA	214.	16.5	17	16.5	Cancer
RBBP5	15.7 ± 9.5	TGCTGGGTGAGAA GGGCTgtggctgcgtttaga GAAGCGTTGGGTAC TGGA	215.	15.5	17	16.5	Epigen etic
RECQL 4	22.1 ± 16.2	TGCGGGACGTGCG GGAGCggctgcaggcgtg ggaGCGCGCGTTCCG ACGGCA	216.	16.5	16	16.5	Cancer

REST	none	TCAGAATACAGTTA TGGCcacccaggtaatggg gCAGTCTTCTGGAG GAGGA	217.	16.5	16	16.5	Epigenetic
RET	5.4 ± 1.8	TGAGTTCTGCCGGC CGCCggctcccgcaggggc caGGGCGAAGTTGG CGCCGA	218.	16.5	17	16.5	Cancer
RNF2	none	TTCTTTATTTCCAG CAATgtctcaggctgtgcag ACAAACGGAATC AACCA	219.	16.5	16	16.5	Epigenetic
RUNX1	25.1 ± 6.9	TTCAGGAGGAAGC GATGGcttcagacagcatat tTGAGTCATTTCTT CGTA	220.	16.5	16	16.5	Epigenetic
SDHB	36.4 ± 19.2	TCTCCTTGAGGCGC CGGTtgcggccacaaccct TGGCGGAGCCTGCC TGCA	221.	16.5	16	16.5	Cancer
SDHC	13.7 ± 3.4	TGTTGCTGAGGTGA CTTCagtgggactgggagtt ggtGCCTGCGGCCCT CCGGA	222.	16.5	19	15.5	Cancer
SDHD	42.0 ± 7.8	TCAGGAACGAGAT GGCGGttctctggaggctga gtGCCGTTTGCGGTG CCCTA	223.	16.5	17	16.5	Cancer
SETDB1	33.5 ± 6.1	TGCAGAGGACAAA AGCATgtcttcccttctgg gTGCAATTGGTTTGG ATGCA	224.	16.5	16	16.5	Epigenetic
SIRT6	43.3 ± 3.1	TTACGCGGCGGGGC TGTCgccgtaocgggacaa gggCAAGTGCGGCC TCCCGGA	225.	16.5	18	16.5	Epigenetic
SMAD2	3.9 ± 1.6	TTTGGTAAGAACAT GTCGtccatcttgcattcac GCCGCCAGTTGTGA AGA	226.	16.5	17	15.5	Cancer
SSI8	31.4 ± 7.9	TGGTGACGGCGGC AACATgtctgtggettgcg ggCCCCGAGGCAGC GAGGCA	227.	16.5	17	16.5	Cancer
SUZ12	13.1 ± 0.4	TGGCGCCTCAGAAG CAcggcgggtggggaggg GGCGGCTCGGGGC CCA	228.	14.5	16	14.5	Epigenetic
TFE3	17.3 ± 2.4	TCATGTCTCATGCG GCCGaaccagctcgggatg gCGTAGAGGCCAGC GCGGA	229.	16.5	16	16.5	Cancer

TGFB2	none	TCGGGGGCTGCTCA GGGGcctgtggccgctgca caTCGTCCTGTGGAC GCGTA	230.	16.5	17	16.5	Cancer
TLX3	none	TTCCGCCCCGCCAG GATGgaggcgcccgccag cgcGCAGACCCCGC ACCCGCA	231.	16.5	17	16.5	Cancer
TP53	19.9 ± 3.6	TTGCCGTCCCAAGC AATGgatgattgatgtgtc CCCGGACGATATTG AACA	232.	16.5	17	16.5	Cancer
TSC2	30.7 ± 22.7	TCCTGGTCCACCAT GGCcaaaccaacaagcaaa gATTCAGGCTTGAA GGAGA	233.	15.5	17	16.5	Cancer
VHL	19.4 ± 1.1	TCTGGATCGCGGAG GGAAtgccccggaggcg gaGAACTGGGACGA GGCCGA	234.	16.5	16	16.5	Cancer
XPA	12.9 ± 2.2	TGGGCCAGAGATG GCGGCggccgacggggt ttgCCGGAGGCGGCG GCTTTA	235.	16.5	16	16.5	Cancer
XPC	31.4 ± 4.2	TGCCCAGACAAGC AACATggctcggaacgc gggccGGCGGGGAG CCGCGGGGA	236.	16.5	19	16.5	Cancer

The abilities of the 96 TALEN pairs to introduce NHEJ-mediated insertion or deletion (indel) mutations at their intended endogenous gene targets were tested in cultured human cells using a slightly modified version of a previously described T7 Endonuclease I (T7EI) assay (Mussolino et al., 2011, Nucleic Acids Res., 39:9283-93; Kim et al., 2009, Genome Res., 19:1279-88). With this T7EI assay, 83 of the 96 TALEN pairs showed evidence of NHEJ-mediated mutagenesis at their intended endogenous gene target sites, an overall success rate of ~86% (Table 7). The efficiencies of TALEN-induced mutagenesis we observed ranged from 2.5% to 55.8% with a mean of 22.5%. To provide molecular confirmation of the mutations we identified by T7EI assay, we sequenced target loci for 11 different TALEN pairs that induced varying efficiencies of mutagenesis (FIGs. 21A-D). As expected, this sequencing revealed indels at the expected target gene sites with frequencies similar to those determined by the T7EI assays.

The nucleotide and amino acid sequences for 14 of the 96 pairs of TALENs targeted to the endogenous human genes in Table 7 are presented below. Each TALEN monomer is presented as follows:

(1) A header with information presented in the format: Gene target_Left or Right monomer_Target DNA site shown 5' to 3'_TALE repeat monomers and 0.5 repeat plasmid used with code as shown in Table 4.

(2) DNA sequence encoding the N-terminal part of the TALE required for activity, the TALE repeat array, the C-terminal 0.5 TALE repeat domain, and the C-terminal 63 amino acids required for activity from a NheI site to a BamHI site. This sequence is present in the "Vector Sequence" plasmid shown below, taking the place of the underlined X's flanked by NheI and BamHI sites

(3) Amino acid sequences the N-terminal part of the TALE required for activity, the TALE repeat array, the C-terminal 0.5 TALE repeat domain, and the C-terminal 63 amino acids required for activity shown from the start of translation (located just 3' to the NheI site and including an N-terminal FLAG epitope tag) to a Gly-Ser sequence (encoded by the BamHI site) that serves as a linker from the TALE repeat array to the FokI cleavage domain.

VECTOR SEQUENCE

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTA
 AGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTGCGTGAGTAGTGCGCGAGCAAAATTTAAGCTACAAC
 AAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATG
 TACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTA
 GTTCATAGCCCATATATGGAGTTCCGCGTTACATAAATTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
 CGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC
 GTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
 CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
 TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCA
 ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTG
 TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCGCCCCATTGACGCAATGGGCGG
 TAGGCGTGACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGC
 TTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCXXXXXXXXXXGGATCCCACTAG
 TCAAAAGTGAAGTGGAGGAGAAGAAATCTGAAGTTCGTCATAAATTGAAATATGTGCCTCATGAATATATT
 GAATTAATTGAATTGCCAGAAATTCCTCAGGATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTAT
 GAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTTACTGTGCG
 GATCTCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGC
 CAAGCAGATGAAATGCAACGATATGTGAAGAAAATCAAACACGAAACAAACATATCAACCCTAATGAATG

GTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAAACT
ACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTT
TTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGG
CGAGATAAACTTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGC
5 GTACCGGTTCATCATCACCATCACCATTGAGTTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTC
CTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT
GAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGC
10 GGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTCT
TCCCTTCCTTCTCGCCACGTTCCCGGCTTCCCGCTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTTC
CGATTAGTGCTTTACGGCACCTCGACCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATC
GCCCTGATAGACGGTTTTTTCGCCCTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCAAA
CTGGAACAACACTCAACCCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGGGGATTTTCGGCCTAT
15 TGGTTAAAAAATGAGCTGATTTAAACAAAAATTTAACCGGAATTAATTCTGTGGAATGTGTGTGCTAGTTAGGG
TGTGGAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCA
GGTGTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC
ATAGTCCCGCCCTAACTCCGCCATCCCGCCCTAACTCCGCCCAGTTCGCCCATTTCTCGCCCCATGG
CTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAG
20 GAGGCTTTTTTGGAGGCCTAGGCTTTTGAAAAAGCTCCCGGGAGCTTGATATCCATTTTCGGATCTGAT
CAGCACGTGTTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAA
ACCATGGCCAAGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCAT
CCCCATCTCTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTCTAGCGACGGCCGATCTTCACTGGTGTCA
ATGTATATCATTTTTACTGGGGGACCTTGTGCGAAGTCTGTTGGTGTGGGCACTGCTGCTGCTGCGGCAGCT
25 GGCAACCTGACTTGTATCGTTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCTGCGGACGGTGTG
ACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAAGGACAGTGATGGACAGCCGACGGCAG
TTGGGATTCTGTAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAAGCACTTCGTGGCCGAGGAGCAGGA
CTGACACGTGCTACGAGATTTGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCC
GGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCCTCGCCCAACCCAACTTGTTT
30 ATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTCACT
GCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTTAGCT
AGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTT
GCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGG
35 GGAGAGGCGGTTTTCGTATTGGGCGCTTTCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTGTTTCG
CTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGG
AAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGCTGGCGTTTTTCC
ATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTAC
40 CGGATACCTGTCCGCTTCTCCTTCGGGAAGCGTGGCGCTTCTCAATGCTCAGCTGTAGGTATCTCA
GTTTCGGTGTAGGTCTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCC
TTATCCGGTAACCTATCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
45 TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAG
AAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTACGGGGTCTGACGCTCAGTGGAAACGAAACTCAC
GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAATAATGAAGT
TTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC

TATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAAACTACGATAC
GGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTA
TCAGCAATAAACCCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCA
GTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
5 TTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTTCAGCTCCGGTTCCCAACGATCA
AGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG
AAGTAAGTTGGCCGCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCAT
CCGTAAGATGCTTTTCTGTGACTGGTGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCG
AGTTGCTCTTGCCCGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCAT
10 TGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCA
CTCGTGACCCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGG
CAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC SEQ ID NO: 244
15

TALE REPEAT SEQUENCES

Target	SEQUENCE	SEQ ID NO:	SEQUENCE	SEQ ID NO:
>APC_Left_TATGTACGCC TCCCTGGG T AL/006/015 /019/025/0 26/012/019 /022/027/0 15/017/022 /027/015/0 19/024/JDS 74/ ('TATGTACG CCTCCCTGGG ' disclosed as SEQ ID NO: 412)	GCTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTATTTCGCAAC AGCAACAGGAGAAAATCAAGCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGACCCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCGAAGCCACGACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAAGTGCACGCTGGCGCAA TCGCTCACCGGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGTCGAACATT GGGGAAAGCAAGCCCTGGAAACCTGCAAAG GTTGTGCGGTCCTTTGTCAAGACCAGGCC TTACACCGGAGCAAGTCTGGCCATTGCAAGC AATGGGGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCTTCCGCTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCTCGAATGGCGGGGTAAGCAGGC GCTGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG	245.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLDT GQLLKIAKRGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNNGGKQALETVQRLLPVLCQAHGLTDPQVVAIANN NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG KQALETVQRLLPVLCQDHGLTDPQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET VQRLLPVLCQAHGLTDPQVVAIANNNGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTDPQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAH GLTDPQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTDPQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNNGGKQALETVQRLLPVLCQAHGLTDPQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRP ALESIVAQLSRPDPALAALTNHDLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	246.

GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCCACGACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAAACAAGCATTGGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTGTGACGCCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGGAACAGTACAGCGCCTGCT		

	GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>APC_Right _TACAGAAGC GGGCAAAGC TAL/006/01 2/016/024/ 026/011/01 9/022/029/ 014/019/02 2/026/011/ 016/024/JD S74/ ('TACAGAAG CGGGCAAAGC ' disclosed as SEQ ID NO: 413)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGCCCCCAAGAAGAAG AGGAAGTGGGCATTACCCGCGGGTACCTAT GGTGGACTTGAGGACATCGGTTATTGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTCGCGCGCTTGGGAGC GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGTCGAACATT GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGCTCCTTTGTCAAGACCACGGCC TTACACCGAGCAAGTCTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTCCCGTGTGTGTC AAGCCCACGGTTTGACGCGCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC	247.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDVT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCDHGLTDPQVVAIASNIGGKQ LETVQRLLPVLCDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCAHGLTPDQVVAIANNNGKQALETVQR LLPVLCAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCDHGLTPDQVVAIANNNGKQALETVQRLLPVL QDHGLTPEQVVAIANNNGKQALETVQRLLPVLCAH GLTPDQVVAIANNNGKQALETVQRLLPVLCAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCAHGLTPDQVVAIAS NIGGKQALETVQRLLPVLCAHGLTPAQVVAIANNNG GKQALETVQRLLPVLCDHGLTPEQVVAIANNNGRP ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	248.

GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGAAAGCAAGCCC TGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGAACATTGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC		

	AAGTGGTCGCCATCGCCAACAACACGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>BRCA1_Lef t_TCCGAAGC TGACAGATGG TAL/007/0 12/019/021 /026/014/0 17/025/029 /011/017/0 21/029/011 /020/024/J DS74/ (TCCGAAGC TGACAGATGG ' disclosed as SEQ ID NO: 414)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGCCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGCGATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAAGT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGCGTCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCTGGCCATTGCATCC CAGGACGCTGGCAAACAGGCTCTTGAGACGCT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC	249.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHEAL VGHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPLQLDT GQLLKIARKGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCDHGLTDPQVVAIANN NGGKQALETVQRLLPVLCDHGLTPEQVVAIASNIGG KQALETVQRLLPVLCDHGLTDPQVVAIASNIGGKQ LETVQRLLPVLCDHGLTPEQVVAIANNNGGKQALET VQRLLPVLCDHGLTDPQVVAIASHDGGKQALETVQR LLPVLCDHGLTPEQVVAIASNIGGKQALETVQRLLP VLCDHGLTDPQVVAIANNNGGKQALETVQRLLPVL CDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCDH GLTDPQVVAIASHDGGKQALETVQRLLPVLCDHGLT PEQVVAIASNIGGKQALETVQRLLPVLCDHGLTDPQ VVAIANNNGGKQALETVQRLLPVLCDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCDHGLTDPQVVAIAS NIGGKQALETVQRLLPVLCDHGLTPEQVVAIANNNG GKQALETVQRLLPVLCDHGLTPEQVVAIANNNGGRP ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGPHAPALIKRTNRRIPERTSHRVAGS	250.

AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCCAATATTGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCCTTTGTCAAGACCAGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCGAATGGCGGCGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCCAATA TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGA ACAATAATGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAAACAA		

	GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAAGAACAACCGGGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>BRCA1_Rig ht_TCAGGTT CCGCCCTAC C_TAL/007/ 011/019/02 4/030/015/ 017/022/02 9/012/017/ 022/027/01 5/016/022/ JDS71/ (TCAGGTT CGCCCCTACC ' disclosed as SEQ ID N: 415)	GCTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTCACCGCGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTGAGG AGCACCCTCGCGCAACACCACGAGCGGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTGGTAAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGAAACCGTGCAAAG GTTGTGCGCGTCCCTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCTGGCCATTGCAAGC AACATCGGTGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT	251.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQ LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKGLPHAPALIKRTNRRIPERTSHRVAGS	252.

GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCAGGCCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGAAAGCAAGCCC TGGAACCGTGCAAAGGTTGTTGCCGGTCTT TGTCAAGACCACGGCCTTACACCGAGCAAGT CGTGGCCATTGCATCCACGACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG		

	TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGAACATTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGGCGCTGGAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCCACGAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGGGGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>BRCA2_Lef t_TTAGACTT AGGTAAGTAA _TAL/010/0 11/019/021 /027/015/0 20/021/029 /014/020/0 21/026/014 /020/021/J DS70/ (TTAGACTT AGGTAAGTAA ' disclosed as SEQ ID NO: 416)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGCGGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGACGAGGCAATTGTAG GGGTCCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGTCAAACGGA GGGGGAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGCTCCTTTGTCAGACACCGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AACATCGGTGGCAACAGGCTCTTGAGACGGT	253.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQD GQLLKIARGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNGGKQALETVQRLLPVLCDHGLTPEQVVAI ASNIGGKQALETVQRLLPVLCAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCAHGLTPAQVVAIASNIGG KQALETVQRLLPVLCDHGLTPDQVVAIASHDGGKQ LETVQRLLPVLCDHGLTPEQVVAIASNGGKQALET VQRLLPVLCAHGLTPDQVVAIASNGGKQALETVQR LLPVLCAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCDHGLTPDQVVAIANNNGGKQALETVQRLLPVL CDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCAH GLTPDQVVAIASNGGKQALETVQRLLPVLCAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IANNGGKQALETVQRLLPVLCAHGLTPDQVVAIAS NGGKQALETVQRLLPVLCAHGLTPAQVVAIASNIG GKQALETVQRLLPVLCDHGLTPEQVVAIASNIGGRP ALESIVAQLSRPDALAALTNHDLVALACLGGRPALD AVKKGLPAPALIKRTNRRIPERTSHRVAGS	254.

TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCCAATATTGGCGGTAAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTTGC CGGTCTTTGTCAAGACCAGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA CGGTGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGAAAGCAAGCCC TGGAACCGGTGCAAAGGTTGTTGCCGGTCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCTCCAATA TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA		

	TTGCAAATAATAACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCTCCAATATTGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGGCCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCCCGATCCCGCCTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATGA TCAAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>BRCA2_Rig ht_TAGTTTG TAGTTCTCCC C_TAL/006/ 014/020/02 5/030/014/ 020/021/02 9/015/020/ 022/030/01 2/017/022/ JDS71/ ('TAGTTTGT AGTTCTCCCC ' disclosed as SEQ ID NO: 417)	GCTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCCTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGCACGAGGCAATTGTAG GGGTGCGTAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGCGCTCACCGGGGGCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCTCGAACATT GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGGTCCTTTGTCAAGACCACGGCC	255.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGSYQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTPDQV VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQ LETVQRLLPVLCQDHGLTPEQVVAIANNNNGGKQALE VQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNNGKQALETVQRLLPVLC QDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNNGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRP	256.

TTACACCGGAGCAAGTCGTGGCCATTGCAAAT AATAACGGTGGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCCAACGGTGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCCTGCACAAGTGGTC GCCATCGCCTCGAATGGCGGGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCGTGAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA CGGTGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGCGCTGGAAA CAGTACAGCGCCTGCTGCCCTGACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGAAAGCAAGCCC TGGAACCGTGCAAAGTTGTTGCCGTCCTT TGTCAAGACCACGGCCTTACACGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGCGCTGGAAACAGTACAG CGCCTGCTGCCGTGACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCCT CAAACGGAGGGGAAAGCAAGCCCTGGAAACC	ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGCATGACGGAGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGCGCTGGAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCCACGAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCCGCGCTCGATGCAGT CAAAAAGGCTCTGCCTCATGCTCCCGCATTGA TCAAAAGAACCAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>ERCC2_Lef t_TCCGGCCG GCGCCATGA_ TAL/007/01 2/019/024/ 027/012/01 9/024/027/ 014/017/02 2/026/015/ 034/JDS70/ ('TCCGGCCG GCGCCATGA' disclosed as SEQ ID NO: 418)	GCTAGCaccatGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTCACCGCGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAAGT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCCGGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC	257.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKRKVG HRCVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQD TQQLLKIARKGGTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQ ALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALE TVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQ RLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN	258.

GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCACGACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGGA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG	NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIG GRPALESIVAQLSRPDPALAALTNHDLVALACLGGRP ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAGCCACGGTCTGACACCGAAC AGGTGGTCGCCATTGCTTCTAACATCGGAGGA CGGCCAGCCTTGAGTCCATCGTAGCCCAATT GTCCAGGCCCGATCCCGGCTTGGCTGCGTTAA CGAATGACCATCTGGTGGCGTTGGCATCTCTT GGTGGACGACCCGCGCTCGATGCAGTCAAAAA GGGTCTGCCTCATGCTCCCGCATTGATCAAAA GAACCAACCGCGGATTCCCGAGAGAACTCC CATCGAGTCGCGGGATCC			
>ERCC2_Rig ht_TAGCGAG CGCGACCC TAL/006/01 4/017/024/ 026/014/01 7/024/027/ 014/016/02 2/027/012/ JDS71/ ('TAGCGAGC GCGACCC' disclosed as SEQ ID NO: 419)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTCACCGCGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTCGCGCGCTTGGGAGC GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCGAAGCCACGCACGAGGCAATTGTAG GGGTGGTAAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGCGCTCACCGGGGCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCGAACATT GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG	259.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ LETVQRLLPVLCQDHGLTPEQVVAIANNGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIANNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIANNGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS HDGGRPALESIVAQLSRPDPALAALTNDHLVALACLG	260.

GTTGTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAA AATAACGGTGGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCAGTCAAGTTGTAGCGATT GCGTCGCATGACGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGSTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTGTC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGGAGCTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTGTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGAACATTGGAG GGAAACAAGCATTGGGAGCTGTCCAACGGCTC CTTCCTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT	GRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	CACATGACGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTCTCTG TCAAGCCACGGGCTGACACCGAACAGGTGG TCGCCATTGCTTCCACGACGGAGGACGGCCA GCCTTGGAGTCCATCGTAGCCCAATTGTCCAG GCCCCATCCCGCTTGGCTGCGTTAACGAATG ACCATCTGGTGGCGTTGGCATGCTTGGTGGA CGACCCGCGCTCGATGCAGTCAAAAAGGTCT GCCTCATGCTCCCGCATTGATCAAAAGAACCA ACCGGCGGATTCCCGAGAGAACTTCCCATCGA GTCGCGGGATCC			
>FANCA_Lef t_TAGGCGCC AAGGCCATGT _TAL/006/0 14/019/022 /029/012/0 17/021/026 /014/019/0 22/027/011 /020/024/J DS78/ (TAGGCGCC AAGGCCATGT ' disclosed as SEQ ID NO: 420)	GCTAGCaccatGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGCACGAGGCAATTGTAG GGGTCCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCGAACATT GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGCGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGCCATTGCAAAT AATAACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAGCCCC	261.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVAI ANNNGKQALETVQRLLPVLCAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCDHGLTPDQVVAIANNNGKQALE TVQRLLPVLCDHGLTPEQVVAIASHDGGKQALETVQR LLPVLCAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVL CDHGLTPEQVVAIANNNGKQALETVQRLLPVLCAH GLTPDQVVAIANNNGKQALETVQRLLPVLCAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCAHGLTPDQVVAIAS NGGKQALETVQRLLPVLCAHGLTPAQVVAIANNNG GKQALETVQRLLPVLCDHGLTPEQVVAIASNGGGRP ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	262.

ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGAAGTACCCAGACCAG GTAGTCGCAATCGCGAACAATAATGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCACGACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGAAGTACCCAGACCAGGTAGTCGC AATCGCGTCGAACATTGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCTTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT		

	GAGACGGTTCAGAGACTTCTCCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAATGGG GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCCGATCCCGCTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>FANCA_Rig ht_TGGCCCG AGGCGGAGTT C_TAL/009/ 014/017/02 2/027/014/ 016/024/02 9/012/019/ 024/026/01 4/020/025/ JDS71/ ('TGGCCCGA GGCGGAGTTC ' disclosed as SEQ ID NO: 421)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACATCGGTTATTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTGGTAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGAACAATAAT GGGGAAAGCAAGCCCTGAAACCGTGCAAAG GTTGTTGCCGCTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCTGGCCATTGCAAAT	263.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGKQALETVQRLLPVLCDHGLTPEQVVAI ANNNGKQALETVQRLLPVLCAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCDHGLTPEQVVAIANNNGKQALET VQRLLPVLCAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCAHGLTPAQVVAIANNNGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGKQALETVQRLLPVL CDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCAH GLTPDQVVAIANNNGKQALETVQRLLPVLCAHGLT PAQVVAIANNNGKQALETVQRLLPVLCDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IANNNNGKQALETVQRLLPVLCAHGLTPDQVVAIAS NNGGKQALETVQRLLPVLCAHGLTPAQVVAIASNNG GKQALETVQRLLPVLCDHGLTPEQVVAIASHDGGR PALD	264.

AATAACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCGTACTGTGCCAG GATCATGGACTGACCCAGACCAGTAGTCGC AATCGCGAACAATAATGGGGGAAAGCAAGCCC TGGAACCGGTGCAAAGGTTGTTGCCGGTCTT TGTCAGACCACGGCCTTACACCGAGCAAGT CGTGGCCATTGCATCCACGACGGTGGCAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAACAACA ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA	AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAATAATAACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCTCGAATGGCGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCCACGAC GGAGGACGGCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>FANCC_Lef t TGAAGGGA CATCACCTTT _TAL/009/0 11/016/024 /029/014/0 16/022/026 /015/017/0 21/027/012 /020/025/J DS78/ (TGAAGGGA CATCACCTTT ' disclosed as SEQ ID NO: 422)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTATTTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGCGATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGAATTGTAG GGGTCCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGAACAATAAT GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG	265.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRCVPMVDLRTLGSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVCKQWSGARALEALLTVAGELRGPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQ LETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG	266.

GTTGTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AACATCGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGAACAATAATGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTGTC CGGTCTTTGTCAAGACCACGGCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTGTGACGCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCGAACATTGGGGGAAAGCAAGCCC TGGAACCGTGCAAAGTTGTTGCCGGTCTT TGTCAAGACCACGGCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGTC CTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCTCCAATA TTGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCTGTGCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT	GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGRP ALESIVAQLSRPDPAALNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGCGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCTCGAATGGCGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCTGCT GCCTGTACTGTGCCAGGATCATGGAAGTACAC CCGAACAGGTGGTCGCCATTGCTTCTAATGGG GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>FANCC_Rig ht_TCTACTG AATCTTGAGC _TAL/007/0 15/016/022 /030/014/0 16/021/030 /012/020/0 25/029/011 /034/JDS71 / ('TCTACTGA ATCTTGAGC' disclosed as SEQ ID NO: 423)	GCTAGCaccatGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGAGGCGCTTGT GGGCGATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGAGTGCACGCTTGGCGCAA TGCGCTCACCAGGGGCCCTTGAACCTGACCC	267.	ASTMDYKDHGDYKDHDIDYKDDDDKMAPKKRKVGI HRCVPMVDLRTLGSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQ LETQRLLPVLCQDHGLTPEQVVAIANNGGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQ LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLP VLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQ VVAIANNGGKQALETVQRLLPVLCQDHGLTPEQVVA	268.

CAGACCAGGTAGTCGCAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AATGGGGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAACCGTGCAAAGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCAAACGGAGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCATCCACGACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG GGAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCGAATG GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG	IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG GRPALESIVAQLSRPDPALAALTNDHLVALACLGGRP ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGA ACAATAATGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTCTGACCCGAAC AGGTGGTCGCCATTGCTTCCCACGAGGAGGA CGGCCAGCCTTGAGTCCATCGTAGCCCAATT GTCCAGGCCCGATCCCGGCTGGCTGCGTTAA CGAATGACCATCTGGTGGCGTTGGCATGTCTT GGTGGACGACCCGCGCTCGATGCAGTCAAAA GGGTCTGCCTCATGCTCCCGCATTGATCAAAA GAACCAACCGCGGATTCCCGAGAGAACTCC CATCGAGTCGCGGGATCC			
>FANCG_Lef t_TCGGCCAC CATGTCCC_T AL/007/014 /019/022/0 27/011/017 /022/026/0 15/019/025 /027/012/J DS71/ ('TCGGCCAC CATGTCCC' disclosed as SEQ ID NO: 424)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTATTTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCCTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTCACGCGCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC	269.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKRKVG HRCVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ANNNGKQALETVQRLLPVLCQAHGLTPDQVVAIANN NGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH GLTPDQVVAIANNNGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS	270.

GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAAT AATAACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCGAACATTGGGGGAAAGCAAGCCC TGGAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCGAATG GCGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG	HDGGRPALESIVAQLSRPDALAALTNDHLVALACLG GRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACACCCGAACAGGTGG TCGCCATTGCTTCCCACGACGGAGGACGGCCA GCCTTGGAGTCCATCGTAGCCCAATTGTCCAG GCCCATCCCGGCTGGCTGCGTTAACGAATG ACCATCTGGTGGCGTTGGCATGTCTTGGTGGA CGACCCGCGCTCGATGCAGTCAAAAAGGCTCT GCCTCATGCTCCCGCATTGATCAAAAGAACCA ACCGGCGGATTCCCGAGAGAACTTCCCATCGA GTCGCGGGATCC			
>FANCG_Rig ht_TCCAGGC AGCTGGAGCC C_TAL/007/ 012/016/02 4/029/012/ 016/024/02 7/015/019/ 024/026/01 4/017/022/ JDS71/ ('TCCAGGCA GCTGGAGCCC ' disclosed as SEQ ID NO: 425)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTGAGG AGCACCGTCGCGCAACACCACGAGCGGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGACCCTGCGGCGCTTGGGAGG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGGTCACATGAC GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGGTCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT	271.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLD GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIANNNNGGKQ LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQAHGLTPAQVVAIANNNNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAH GLTPDQVVAIANNNNGGKQALETVQRLLPVLCQAHGLT PAQVVAIANNNNGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA IANNNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGRP ALESIVAQLSRPDALAALTNHDLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	272.

TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGAACAATAATGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCACGACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAACCGGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAACAACA ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGCCA		

	TTGCAAATAATAACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGCATGACGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCTGCT GCCTGTACTGTCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCACGAC GGAGGACGGCCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAGGGTCTGCCTCATGCTCCCGATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>JAK2_Left _TCTGAAAA GACTCTGCA TAL/007/01 5/019/021/ 026/011/01 6/021/029/ 011/017/02 5/027/015/ 019/022/JD S70/ ('TCTGAAAA AGACTCTGCA ' disclosed as SEQ ID NO: 426)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCCGAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGGCGGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAATAACCAAGATATGATTGCGGC CCTGCCGAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCCGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCCTCACCGGGGGCCCCCTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC GGGGGAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGCGTCTTTGTCAAGACCAGGCC	273.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKGVI HRCVPMVDLRTLGYSQLQKEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQDLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDG GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP	274.

TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AATGGGGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCCAATATTGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGTAGTCGC AATCGCGAACAATAATGGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTGCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCGAATG GCGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGGAAAGCAAGCCCTGGAAACC	ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGCGCTGGAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGCTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>JAK2 Right TCCATTTC TGTCATCGTA _TAL/007/0 12/016/025 /030/015/0 17/025/029 /015/017/0 21/030/012 /019/025/J DS70/ ('TCCATTTC TGTCATCGTA ' disclosed as SEQ ID NO: 427)	GGTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTCACCGCGGGTACCTAT GGTGGACTTGAGGACACTCGGTATTTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC	275.	ASTMDYKDHGDYKDHDIYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQD GQLLKIATRGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQ LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNGGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN	276.

GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCTCGAATGGCGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCTGCACAAGTGGTCGCCATCGC CTCGAATGGCGCGGTAAGCAGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGTGGCAAAC AGGCTCTTGAGACGTTTCAAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCTCCAATA TTGGCGGTAAGCAGCGCTGGAAACAGTACAG CGCTGCTGCCTGTACTGTGCCAGGATCATGG	NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGRP ALESIVAQLSRDPALAAALNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	ACTGACCCAGACCAGGTAGTCGCAATCGCGT CAAACGGAGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGGCGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>KRAS_Left _TGAAATGA CTGAATATA TAL/009/01 1/016/021/ 026/015/01 9/021/027/ 015/019/02 1/026/015/ 016/025/JD S70/ ('TGAAATG ACTGAATATA ' disclosed as SEQ ID	GCTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCCTCGCGCAACACCACGAGGCGCTTGT GGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTCGGTAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGAGTGCACGCGCTGGCGCAA	277.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKGVI HRGVPMVDLRTLGYSQQQEKIKPKVRSTVAQHHEAL VGHGFTAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVKQWNGARALEALLTVAGELRGPPLQDLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ LETVQRLLPVLCQDHGLTPEQVVAIASNNGGKQALET VQRLLPVLCQAHGLTPDQVVAIANNGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAH GLTPDQVVAIANNGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ	278.

NO: 428)	TGCCTCACCAGGGCCCCCTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGAACAATAAT GGGGAAAGCAAGCCCTGGAAACCGTGCAAAG GTTGTGCGCGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AACATCGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGAACATTGGAGGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCCAATATTGGCGGTAAAGAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAAGCAGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCCAATA	VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IASNGGKQALETVQRLLPVLCAHGLTPDQVVAIAS NIGGKQALETVQRLLPVLCAHGLTPAQVVAIASNGG GKQALETVQRLLPVLCDHGLTPEQVVAIASNIGGRP ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPTSHRVAGS	

	TTGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAATGGGGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTGAACATTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGCCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTGGTGGAGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>KRAS_Righ t_TTGCTAC GCCACCAGC TAL/010/01 4/017/022/ 030/011/01 7/024/027/ 012/016/02 2/027/011/ 034/JDS71/ 'TTGCCTAC GCCACCAGC' disclosed as SEQ ID	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGACCCCTCGGCGGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGCTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC	279.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNGGKQALETVQRLLPVLCQDHGLTPEQVVAI ANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGKQALE LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIANNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH	280.

NO: 429)	AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCAAACGGA GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG GTTGTTGCCGCTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAAAT AATAACGGTGGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGGA CAGTACAGCGCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAACCGTGCAAAGGTTGTTGCCGCTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCATCCACGACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGTCGAACATTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC	GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGKQALETVQRLLPVLCQDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDG GRPALESIVAQLSRPDPALAAALNDHLVALACLGGRP ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS	

	CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTCTGACACCGAAC AGGTGGTCGCCATTGCTTCCCACGACGGAGGA CGGCCAGCCTTGGAGTCCATCGTAGCCCAATT GTCCAGGCCGATCCCGGTTGGCTGCGTTAA CGAATGACCATCTGGTGGCGTTGGCATGTCTT GGTGGACGACCCGCGCTCGATGCAGTCAAAA GGGTCTGCCTCATGTCCCGCATTGATCAAAA GAACCAACCGGGGATTCCCGAGAGAACTTCC CATCGAGTCGCGGGATCC			
>MYC_Left TGCTTAGACG CTGGATTT T AL/009/012 /020/025/0 26/014/016 /022/029/0 12/020/024 /029/011/0 20/025/JDS 78/ 'TGCTTAGA CGCTGATTT ' disclosed	GCTAGCaccatGGACTACAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCCTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCGAAGCCACGACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGGAAGAGAGGGGGAGTA	281.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKRKVG HRGVPMVDLRTLGSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPEQVVAIANNNGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGKQALETVQRLLPVLC QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLT	282.

as SEQ ID NO: 430)	ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TCCGCTCACCGGGGCCCCCTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGAACAATAAT GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG GTTGTTGCCGGTCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CAGCAGCGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCCAACGGTGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCGAATGGCGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGAACAATAATGGGGGAAAGCAAGCCC TGGAACCCGTGCAAAGGTTGTTGCCGGTCTT TGTCAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCATCCACGACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC	PAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQ VVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG GKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGRP ALESIVAQLSRPDPALAAALNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	GCCTGCACAAGTGGTCGCCATGCCAACACA ACGGCGGTAAGCAGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGA ACAATAATGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT AAGCAGGCGCTGGAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAATGGG GGAGGACGGCCAGCCTTGGAGTCCATCCTAGC CCAATTGTCCAGGCCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>MYC_Right _TTCGGTGCT TACCTGGTT TAL/010/01 2/019/024/ 030/014/01 7/025/030/ 011/017/02 2/030/014/ 019/025/JD S78/ {'TTCGGTGC TTACCTGGTT	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCCCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT	283.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPQLD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNGGKQALETVQRLLPVLCDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCDHGLTPDQVVAIASNGGKQALE LETVQRLLPVLCDHGLTPEQVVAIANNNGGKQALET VQRLLPVLCAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCAHGLTPAQVVAIASNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNGGKQALETVQRLLPVL	284.

disclosed as SEQ ID NO: 431)	TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCAAACGGA GGGGAAAGCAAGCCCTGAAACCGTGCAAAG GTTGTTGCCGGTCCCTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCGAATGGCGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCAAACGGAGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCCTT TGTCAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG	QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA IANNNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG GKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGGRP ALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	

	GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CAAACGGAGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGCCA TTGCAAATAATAACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCCGTCAC AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT AAGCAGGCGCTGGAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAATGGG GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>PTEN_Left _TCCCAGACA TGACAGCC_T AL/007/012 /017/021/0 29/011/017 /021/030/0 14/016/022 /026/014/0 32/JDS71/ ('TCCCAGAC	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCTG CGCTTTCACAGCACCTCGCGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG	285.	ASTMDYKDHGDYKDHIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQ LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR	286.

ATGACAGCC' disclosed as SEQ ID NO: 432)	GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG GTTGTTGCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTCCCGTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCTCCAATATTGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCGTGAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGGGAACAATAATGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTGAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGAAG CAGTACAGCGCCTGCTGCCGTGACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCAAACGGAGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA	LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVL QDHGLTPEQVVAIANNNIGGKQALETVQRLLPVLCAH GLTPDQVVAIASNIGGKQALETVQRLLPVLCAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IANNNIGGKQALETVQRLLPVLCAHGLTPDQVVAIAS HDGGKQALETVQRLLPVLCAHGLTPEQVVAIASHDG GRPALESIVAQLSRPDPALAALNDHLVALACLGGRP ALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGAACATTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGCTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAATAATAACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGCATGACGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTCTGACACCGAAC AGGTGGTCGCCATTGCTTCCACGACGGAGGA CGGCCAGCCTTGAGTCCATCGTAGCCCAATT GTCCAGGCCCGATCCCGCTTGGCTGCGTTAA CGAATGACCATCTGGTGGCGTTGGCATGTCTT GGTGGACGACCCGCGCTCGATGCAGTCAAAAA GGGTCTGCCTCATGCTCCCGCATTGATCAAAA GAACCAACCGGCGGATTCCCGAGAGAACTTCC CATCGAGTCGCGGGATCC			
>PTEN_Righ t_TCCTTTTG TTTCTGCTAA _TAL/007/0 12/020/025 /030/015/0 19/025/030 /015/017/0 25/029/012 /020/021/J DS70/ ('TCCTTTTG	GCTAGCaccatGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGAGCGGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGACCCCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCGAAGCCACGACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGGAGCA	287.	ASTMDYKDHGDYKDHDIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPLQLDT GQLLKIAKRGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALE LETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALET VQRLLPVLCQAHGLTPDQVVAIANNGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLP	288.

TTTCTGCTAA , disclosed as SEQ ID NO: 433)	CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGGCGCTCACCGGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCCGAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAG GTTGTGGCGGTCCTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCCAACGGTGGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCTCGAATGGCGGCGGTAAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCAAACGGAGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CTCGAATGGCGGCGGTAAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCAAACGGAGGGGGAAGCAAGCCC TGGAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA	VLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGKQALETVQRLLPVLCQAHGLT PAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ VVAIANNNNGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGKQALETVQRLLPVLCQAHGLTPEQVVAIAS NNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIG GKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGRP ALESIVAQLSRPDALAALTNHDLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	

	TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCGAATG GCGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGA ACAATAATGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGCTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCCAACGGTGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCTCCAATATTGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAACATC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGCTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>TP53 Left _TTGCCGTCC CAAGCAATG_ TAL/010/01 4/017/022/ 029/015/01 7/022/027/ 011/016/02 4/027/011/ 016/025/JD	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCCTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC	289.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGI HRCVPMVDLRTLGSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNGGKQALETVQRLLPVLCDHGLTPEQVVAI ANNNGGKQALETVQRLLPVLCAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCDHGLTPDQVVAIANNNGGKQA LETVQRLLPVLCDHGLTPEQVVAIASNGGKQALET	290.

S74/ ('TTGCCGTC CCAAGCAATG ' disclosed as SEQ ID NO: 434)	CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCCGGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCCGAATCGCGTCAAACGGA GGGGAAAGCAAAGCCCTGAAACCGTGCAAAG GTTGTGCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAAT AATAACGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGAACAATAATGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAATGGGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGCA TGACGGAGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGGA CAGTACAGCGCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTGCGGTCCTT TGTCAAGACCACGGCCTTACACGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAAC	VQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAH GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLT PAQVVAIANNGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS NIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGG GKQALETVQRLLPVLCQDHGLTPEQVVAIANNGGRP ALESIVAQLSRDPALAAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS	

	AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGAACATTGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAACAACA ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCTGTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCAAGCAACATCGGTGGCAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGAACATTGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCTCGAATGGCGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCTGCT GCCTGTACTGTGCCAGGATCATGGAAGTACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>TP53_Righ t_TGTTCAAT ATCGTCCGGG _TAL/009/0 15/020/022 /026/011/0 20/021/030 /012/019/0	GCTAGCaccATGGACTACAAAGACCATGACGS TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGGCGGCTTGT GGGCATGGCTTCACTCATGCGCATATTGTGCG	291.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGGKQALETVQRLLPVLQDQDHLTPAQVVAI ASNGGKQALETVQRLLPVLQAHGLTPDQVVAIASN GGGKQALETVQRLLPVLQAHGLTPAQVVAIASHDGG	292.

25/027/012 /019/024/J DS74/ ('TGTTCAAT ATCGTCCGGG ' disclosed as SEQ ID NO: 435)	CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCGAAGCCACGCACGAGGCAATTGTAG GGGTCCGTAACAGTGGTCCGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGAACAATAAT GGGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AATGGGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCCAACGGTGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCCAA CGGTGGAGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCAAACGAGGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTTGTTGCCGTCCTT	KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC QDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAH GLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLT PAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNG GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGRP ALESIVAQLSRPDALAALTNDHLVALACLGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

	TGTC AAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCATCCCACGACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCTCGAATG GCGGCGGTAAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGTCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGCTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>XPA_Left_ TGGGCCAGAG ATGGCGGC_T AL/009/014 /019/022/0 27/011/019	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCAATCACCGCGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAATCAAGCCTAAGGTCAGG	293.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKRKVG HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAI	294.

/021/029/0 11/020/024 /029/012/0 19/024/JDS 71/ ('TGGGCCAG AGATGGCGGC ' disclosed as SEQ ID NO: 436)	AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCCATATTGTGCG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGACGAGGCAATTGTAG GGGTCCGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCCGGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGGCCCCCTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGAACAATAAT GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTTGCGGTCCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAT AATAACCGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGAATAACAATGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTCTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCTGCACAAGTGGTCGCCATCGC CTCCAATATTGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC	ANNNGGKQALETVQRLLPVLCAHGLTPDQVVAIANN NGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCDHGLTPEQVVAIASNIGGKQALET VQRLLPVLCAHGLTPDQVVAIANNNGGKQALETVQR LLPVLCAHGLTPAQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVL QDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCAH GLTPDQVVAIASNIGGKQALETVQRLLPVLCAHGLT PAQVVAIANNNGGKQALETVQRLLPVLCDHGLTPDQ VVAIANNNGGKQALETVQRLLPVLCDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCAHGLTPAQVVAIANNNG GKQALETVQRLLPVLCDHGLTPEQVVAIASHDGGRP ALESIVAQLSRPDALAALTDNHLVALACLGGRPALD AVKGLPHAPALIKRTNRRIPERTSHRVAGS	

	AATCGCGAACAATAATGGGGGAAAGCAAGCCC TGGAAACCGTGCAAAGGTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCCAACGGTGGAG GGAAACAAGCATTTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAACAACA ACGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAAGTAGTCGCAATCGCGA ACAATAATGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTGTTGCCGGTCCTTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCCACGAC GGAGGACGGCCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>XPA_Right _TAAAGCCGC CGCCTCCGG _TAL/006/01	GCTAGCaccATGGACTACAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT	295.	ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKVGI HRGVPMVDLRTLGYSQLQEKIKPKVRSTVAQHHEAL VHGFTAHIVALSQLHPAALGTVAVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT	296.

1/016/024/ 027/012/01 9/022/027/ 014/017/02 2/030/012/ 017/024/JD S74/ ('TAAAGCCG CCGCTCCGG ' disclosed as SEQ ID NO: 437)	GGTGGACTTGAGGACACTCGGTTATTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTCG CGCTTTCACAGCACCTCGGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTCGGTAAACAGTGGTCGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGGCCCTTGAACCTGACCC CAGACCAGGTAGTCGCAATCGCGTCGAACATT GGGGAAAGCAAGCCCTGGAACCGTGCAAAG GTTGTGCGCGTCCCTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCAAGC AACATCGGTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT CGCTCGAACATTGGAGGAAACAAGCATTGGA GACTGTCCAACGGCTCCTTCCCGTGTGTGTC AAGCCCACGGTTTGACGCGCTGCACAAGTGGTC GCCATCGCCAACAACAACGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCACATGACGGGGGAAA GCAAGCCCTGGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCACGACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA CAATGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGGA	GQLLKIARGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAI ASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN IGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGG KQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALET VQRLLPVLCQAHGLTPDQVVAIANNNGKQALETVQR LLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC QDHGLTPEQVVAIANNNGKQALETVQRLLPVLCQAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNG GKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGRP ALESIVAQLSRPDALAALTNDHLVALACLGGRPALD AVKKGLPHAPALIKRINRRIPERTSHRVAGS

	CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAACCGGTGCAAAGGTTGTTGCCGGTCCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAATAATAACGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTTGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CAAACGGAGGGGGAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCCTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGCATGACGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAACAACAACGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTAATAATAAC GGAGGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCGATCCCGGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			
>XPC_Left_ TGCCCAGACA	GCTAGCaccatGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG	297.	ASTMDYKDHGDYKDHDIDYKDDDDKMAPKKRKVGI HRCVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL	298.

AGCAACAT T AL/009/012 /017/022/0 26/014/016 /022/026/0 11/019/022 /026/011/0 17/021/JDS 78/ ('TGCCCGA CAAGCAACAT ' disclosed as SEQ ID NO: 438)	ATGACGATGACAAGATGGCCCCAAGAAGAAG AGGAAGGTGGGCATTACCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTATTTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCAGGAGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTGCGGCGCTTGGGACG GTGGCTGTCAAATACCAAGATATGATTGCGGC CCTGCCCCAAGCCACGCACGAGGCAATTGTAG GGGTGCGTAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGAGTA ACAGCGGTAGAGGCGAGTGCACGCCCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACTGACCC CAGACCAGGTAGTCGCAATCGCGAACAATAAT GGGGAAAGCAAGCCCTGAAACCGTGCAAAG GTTGTTGCGGTCCTTTGTCAAGACCAGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CACGACGTTGGCAAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGGAGGGAACAAGCATTGGA GACTGTCCAACGGGCTCCTTCCCGTGTGTGTC AAGCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGGTAAGCAGGC GCTGGAAACAGTACAGCGCCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCAG GTAGTCGCAATCGCGTCGAACATTGGGGGAAA GCAAGCCCTGAAACCGTGCAAAGGTTGTTGC CGGTCCTTTGTCAAGACCAGGCTTACACCG GAGCAAGTCGTGGCCATTGCAAATAATAACGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGTCGAA CATTGGAGGGAAACAAGCATTGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC	VGHGFTHAHIVALSQHPAALGTAVKYQDMIALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQD GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLT VQV VAIANNNGKQALETVQRLLPVLCDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCDHGLTPEQVVAIASH DGGKQALETVQRLLPVLCDHGLTPEQVVAIASHDGG KQALETVQRLLPVLCDHGLTPEQVVAIASNIGGKQ LETVQRLLPVLCDHGLTPEQVVAIANNNGKQALET VQRLLPVLCDHGLTPEQVVAIASNIGGKQALETVQR LLPVLCDHGLTPEQVVAIASHDGGKQALETVQRLLP VLCQDGLTPEQVVAIASNIGGKQALETVQRLLPVL CDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCDH GLTPEQVVAIASNIGGKQALETVQRLLPVLCDHGLT GLTPEQVVAIASNIGGKQALETVQRLLPVLCDHGLT PAQVVAIASHDGGKQALETVQRLLPVLCDHGLTPEQ VVAIASNIGGKQALETVQRLLPVLCDHGLTPEQVVA IASNIGGKQALETVQRLLPVLCDHGLTPEQVVAIAS HDGGKQALETVQRLLPVLCDHGLTPEQVVAIASNIG GKQALETVQRLLPVLCDHGLTPEQVVAIASNIGGRP ALESIVAQLSRDPALAAALNDHLVALACLGGRPALD AVKKGLPHAPALIKRTNRRIPERTSHRVAGS

GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAGCCATGATGGCGGTAAGCAGGCGCTGGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCGAACATTGGGGGAAAGCAAGCCC TGAAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAACATCGGTGGCAAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCCA GTTCTCTGTCAAGCCCACGGGCTGACTCCGA TCAAGTTGTAGCGATTGCGAATAACAATGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGGAAACAGTACAG CGCCTGCTGCCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CGAACATTGGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGCCA TTGCAAGCAACATCGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGTCGCATGACGGAGGGAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCTCCAATATTGGCGGT AAGCAGGCGCTGGAAACAGTACAGCGCCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCTAATGGG GGAGACGGCCAGCCTTGGAGTCCATCGTAGC CCAATTGTCCAGGCCCCGATCCCGGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTTGGTGGACGACCCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA TCAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC		

>XPC_Right _TCCCCGCGG CTCCCGGCC TAL/007/01 2/017/022/ 029/012/01 9/024/027/ 015/017/02 2/027/012/ 019/022/JD S71/ ('TCCCCGCG GCTCCCGGCC ' disclosed as SEQ ID NO: 439)	GCTAGCaccATGGACTACAAAGACCATGACGG TGATTATAAAGATCATGACATCGATTACAAGG ATGACGATGACAAGATGGCCCCCAAGAAGAAG AGGAAGGTGGGCATTACCCGCGGGGTACCTAT GGTGGACTTGAGGACACTCGGTTATTTCGCAAC AGCAACAGGAGAAAAATCAAGCCTAAGGTCAGG AGCACCGTCGCGCAACACCACGAGGCGCTTGT GGGGCATGGCTTCACTCATGCGCATATTGTGCG CGCTTTCACAGCACCTCGCGGCGCTTGGGACG GTGGCTGTCAATACCAAGATATGATTGCGGC CCTGCCCCGAAGCCACGACGAGGCAATTGTAG GGGTGCGTAAACAGTGGTGGGAGCGCGAGCA CTTGAGGCGCTGCTGACTGTGGCGGTGAGCT TAGGGGGCCTCCGCTCCAGCTCGACACCGGGC AGCTGCTGAAGATCGCGAAGAGAGGGGGAGTA ACAGCGGTAGAGGCAGTGCACGCTGGCGCAA TGCGCTCACCGGGGCCCCCTTGAACCTGACCC CAGACCAGGTAGTCGAATCGCGTCACATGAC GGGGGAAAGCAAGCCCTGGAACCGTGCAAG GTTGTTGCCGGTCTTTGTCAAGACCACGGCC TTACACCGGAGCAAGTCGTGGCCATTGCATCC CAGGACGGTGGCAACAGGCTCTTGAGACGGT TCAGAGACTTCTCCAGTTCTCTGTCAAGCCC ACGGGCTGACTCCCGATCAAGTTGTAGCGATT GCGTCGCATGACGAGGGAACAAGCATTGGA GACTGTCCAACGGCTCTTCCCGTGTGTGTGTC AAGCCCACGGTTTGACGCTGCACAAGTGGTC GCCATCGCCAGCCATGATGGCGTAAGCAGGC GCTGGAACAGTACAGCGCTGCTGCCTGTAC TGTGCCAGGATCATGGACTGACCCAGACCA GTAGTCGCAATCGGCAACAATAATGGGGGAAA GCAAGCCCTGGAACCGTGCAAAGGTTGTTGC CGGTCTTTGTCAAGACCACGGCCTTACACCG GAGCAAGTCGTGGCCATTGCATCCACGACGG TGGCAACACAGGCTCTTGAGACGGTTCAGAGAC TTCTCCAGTTCTCTGTCAAGCCCACGGGCTG ACTCCCGATCAAGTTGTAGCGATTGCGAATAA	299. ASTMDYKDHDGDYKDHDIDYKDDDDKMAPKKRKRKVG HRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGTHAHIVALSQHPAALGTAVVKYQDMIAALPEA THEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDT GQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQV VAIASHDGGKQALETVQRLLPVLCDHGLTPEQVVAI ASHDGGKQALETVQRLLPVLCAHGLTPDQVVAIASH DGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDGG KQALETVQRLLPVLCDHGLTPDQVVAIANNNGGKQA LETVQRLLPVLCDHGLTPEQVVAIASHDGGKQALET VQRLLPVLCAHGLTPDQVVAIANNNGGKQALETVQR LLPVLCAHGLTPAQVVAIANNNGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVL CDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCAH GLTPDQVVAIASHDGGKQALETVQRLLPVLCAHGLT PAQVVAIASHDGGKQALETVQRLLPVLCDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCDHGLTPEQVVA IASHDGGKQALETVQRLLPVLCAHGLTPDQVVAIAN NNGGKQALETVQRLLPVLCAHGLTPAQVVAIASHDG GKQALETVQRLLPVLCDHGLTPEQVVAIASHDGGRP ALESIVAQLSRPDALAALTDHLVALACLGGRPALD AVKKGLPHAPALIKRTRRIPERTSHRVAGS	300.

CAATGGAGGGAAACAAGCATTGGAGACTGTCC AACGGCTCCTTCCCGTGTGTGTCAAGCCAC GGTTTGACGCCTGCACAAGTGGTCGCCATCGC CAACAACAACGGCGGTAAGCAGGCGCTGAAA CAGTACAGCGCCTGCTGCCTGTACTGTGCCAG GATCATGGACTGACCCAGACCAGGTAGTCGC AATCGCGTCACATGACGGGGAAAGCAAGCCC TGGAACCGTGCAAAGGTTGTTGCCGGTCCTT TGTCAAGACCACGGCCTTACACCGGAGCAAGT CGTGGCCATTGCAAGCAATGGGGGTGGCAAC AGGCTCTTGAGACGGTTCAGAGACTTCTCCA GTTCTCTGTCAAGCCACGGGCTGACTCCCGA TCAAGTTGTAGCGATTGCGTCGCATGACGGAG GGAAACAAGCATTGGAGACTGTCCAACGGCTC CTTCCCGTGTGTGTCAAGCCACGGTTTGAC GCCTGCACAAGTGGTCGCCATCGCCAGCCATG ATGGCGGTAAGCAGGCGCTGAAACAGTACAG CGCTGTGCTGCTGTACTGTGCCAGGATCATGG ACTGACCCAGACCAGGTAGTCGCAATCGCGT CACATGACGGGGAAAGCAAGCCCTGGAAACC GTGCAAAGGTTGTTGCCGGTCCTTGTCAAGA CCACGGCCTTACACCGGAGCAAGTCGTGGCCA TTGCATCCCACGACGGTGGCAAACAGGCTCTT GAGACGGTTCAGAGACTTCTCCAGTTCTCTG TCAAGCCACGGGCTGACTCCCGATCAAGTTG TAGCGATTGCGAATAACAATGGAGGGAAACAA GCATTGGAGACTGTCCAACGGCTCCTTCCCGT GTTGTGTCAAGCCACGGTTTGACGCCTGCAC AAGTGGTCGCCATCGCCAGCCATGATGGCGGT AAGCAGGCGCTGAAACAGTACAGCGCTGCT GCCTGTACTGTGCCAGGATCATGGACTGACAC CCGAACAGGTGGTCGCCATTGCTTCCCACGAC GGAGGACGGCCAGCCTTGAGTCCATCGTAGC CCAATTGTCCAGGCCGATCCCGCGTTGGCTG CGTTAACGAATGACCATCTGGTGGCGTTGGCA TGTCTGGTGGACGACCGCGCTCGATGCAGT CAAAAAGGGTCTGCCTCATGCTCCCGCATTGA		

	TCAAAAGAACCAACCGGCGGATTCCCGAGAGA ACTTCCCATCGAGTCGCGGGATCC			

OTHER EMBODIMENTS

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A process comprising:

(a) providing a first nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs: 58-62;

(b) contacting the first nucleic acid with a first enzyme, wherein the first enzyme creates a first ligatable end;

(c) providing a second nucleic acid comprising a β unit, a γ unit, a δ unit, and a ϵ unit, wherein the β unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ϵ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 78-82;

(d) contacting the second nucleic acid with a second enzyme, wherein the second enzyme creates a second ligatable end, and wherein the first and second ligatable ends are compatible; and

(e) ligating the first and second nucleic acids through the first and second ligatable ends to produce a first ligated nucleic acid,

wherein the first ligated nucleic acid is linked to a solid support.

2. The process of claim 1, wherein the first and second enzymes are a first and second restriction endonuclease, wherein the first restriction endonuclease cleaves at a site within the first nucleic acid and creates a first cut end, and the second restriction endonuclease cleaves at a site within the second nucleic acid and creates a second cut end, and wherein the first and second ligatable ends are the first and second cut ends.

3. The process of claim 2, wherein the first ligated nucleic acid does not comprise a restriction site recognized by the first restriction endonuclease.

4. The process of claim 1 or 2, further comprising:

(f) contacting the first ligated nucleic acid with a third enzyme, wherein the third enzyme creates a third ligatable end;

(g) providing a third nucleic acid comprising a β unit, a γ unit, a δ unit, and a ϵ unit, wherein the β unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72, the δ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ϵ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 78-82;

(h) contacting the third nucleic acid with a fourth enzyme, wherein the fourth enzyme creates a fourth ligatable end, and wherein the third and fourth ligatable ends are compatible;

(i) ligating the first ligated and third nucleic acids through the third and fourth ligatable ends to produce a second ligated nucleic acid linked to the solid support; and

(j) optionally repeating steps (f)-(i), wherein the second ligated nucleic acid produced in step (i) is treated as the first nucleic acid in step (f) during the repeating process.

5. The process of claim 4, wherein the third and fourth enzymes are a third and fourth restriction endonuclease, wherein the third restriction endonuclease cleaves at a site within the first ligated nucleic acid and creates a third cut end, and the fourth restriction endonuclease cleaves at a site within the third nucleic acid and creates a fourth cut end, and wherein the third and fourth ligatable ends are the third and fourth cut ends.

6. The process of claim 5, wherein the ligated nucleic acid does not comprise a restriction site recognized by the first endonuclease, and wherein the first and third restriction endonucleases are the same.

7. The process of claim 5 or 6, wherein the second and fourth restriction endonucleases are the same.

8. The process of claim 4 or 5, further comprising:

(k) contacting the second ligated nucleic acid with a fifth enzyme, wherein the fifth enzyme creates a fifth ligatable end;

(l) providing a fourth nucleic acid, wherein the fourth nucleic acid is selected from one of the following nucleic acids:

- i. a nucleic acid comprising a β unit,
- ii. a nucleic acid comprising a β unit and a γ' unit, or
- iii. a nucleic acid comprising a β unit, a γ unit, and a δ unit,

wherein the β unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, the γ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72, the γ' unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 83-87, and the δ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77;

(m) contacting the fourth nucleic acid with a sixth enzyme, wherein the sixth enzyme creates a sixth ligatable end, and wherein the fifth and sixth ligatable ends are compatible; and

(n) ligating the second ligated and fourth nucleic acids through the fifth and sixth ligatable ends to produce a third ligated nucleic acid linked to the solid support.

9. The process of claim 8, wherein the fifth and sixth enzymes are a fifth and sixth restriction endonuclease, wherein the fifth restriction endonuclease cleaves at a site within the second ligated nucleic acid and creates a fifth cut end, and the sixth restriction endonuclease cleaves at a site within the fourth nucleic acid and creates a sixth cut end, and wherein the fifth and sixth ligatable ends are the fifth and sixth cut ends.

10. The process of claim 9, wherein the second ligated nucleic acid does not comprise a restriction site recognized by the first endonuclease, and wherein the first, third, and fifth restriction endonucleases are the same.

11. The process of claim 9 or 10, wherein the second, fourth, and sixth restriction endonucleases are the same.

12. The process of any one of claims 1-11, wherein the first and second ligatable ends each comprise an overhang of 1-10 nucleotides.

13. The process of any one of claims 1-11, wherein the first enzyme is a type IIS restriction endonuclease.
14. The process of any one of claims 1-3, further comprising unlinking the first ligated nucleic acid from the solid support and inserting the first ligated nucleic acid into a vector.
15. The process of any one of claims 4-7, further comprising unlinking the second ligated nucleic acid from the solid support and inserting the second ligated nucleic acid into a vector.
16. The process of any one of claims 8-11, further comprising unlinking the third ligated nucleic acid from the solid support and inserting the third ligated nucleic acid into a vector.
17. The processes of claim 4 or 5, further comprising
 - (k) contacting the second ligated nucleic acid with a fifth enzyme, wherein the fifth enzyme creates a fifth ligatable end;
 - (l) providing a fourth nucleic acid comprising a β unit and a γ unit, wherein the β unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63-67, and the γ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 68-72;
 - (m) contacting the fourth nucleic acid with a sixth enzyme, wherein the sixth enzyme creates a sixth ligatable end, and wherein the fifth and sixth ligatable ends are compatible;
 - (n) ligating the second ligated nucleic acid and the fourth nucleic acid through the fifth and sixth ligatable ends to produce a third ligated nucleic acid linked to the solid support.
18. The process of claim 17, further comprising
 - (o) contacting the third ligated nucleic acid with a seventh enzyme, wherein the seventh enzyme creates a seventh ligatable end;
 - (p) providing a fifth nucleic acid comprising a δ unit and a ϵ' unit, wherein the δ unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 73-77, and the ϵ' unit has a nucleotide sequence selected from the group consisting of SEQ ID NOs: 88-92;
 - (q) contacting the fifth nucleic acid with an eighth enzyme, wherein the eighth enzyme creates an eighth ligatable end, and wherein the seventh and eighth ligatable ends are compatible;

(r) ligating the third ligated nucleic acid and the fifth nucleic acid through the seventh and eighth ligatable ends to produce a fourth ligated nucleic acid linked to the solid support.

19. The process of claim 18, further comprising unlinking the fourth ligated nucleic acid from the solid support and inserting the fourth ligated nucleic acid into a vector.

20. The process of any one of claims 14-16 and 19, wherein the vector is an expression vector.

21. The process of claim 20, wherein the expression vector includes a sequence encoding an effector domain, and wherein the first, second, or third ligated nucleic acid is inserted into the vector such that the vector comprises a sequence encoding a fusion protein of the polypeptide and the effector domain.

22. The process of claim 21, wherein the effector domain is a nuclease domain.

23. The process of any one of claims 20-22, further comprising inserting the expression vector into a cell.

24. The process of claim 23, further comprising expressing the polypeptide or fusion protein.

25. The process of claim 24, further comprising purifying the polypeptide or fusion protein.

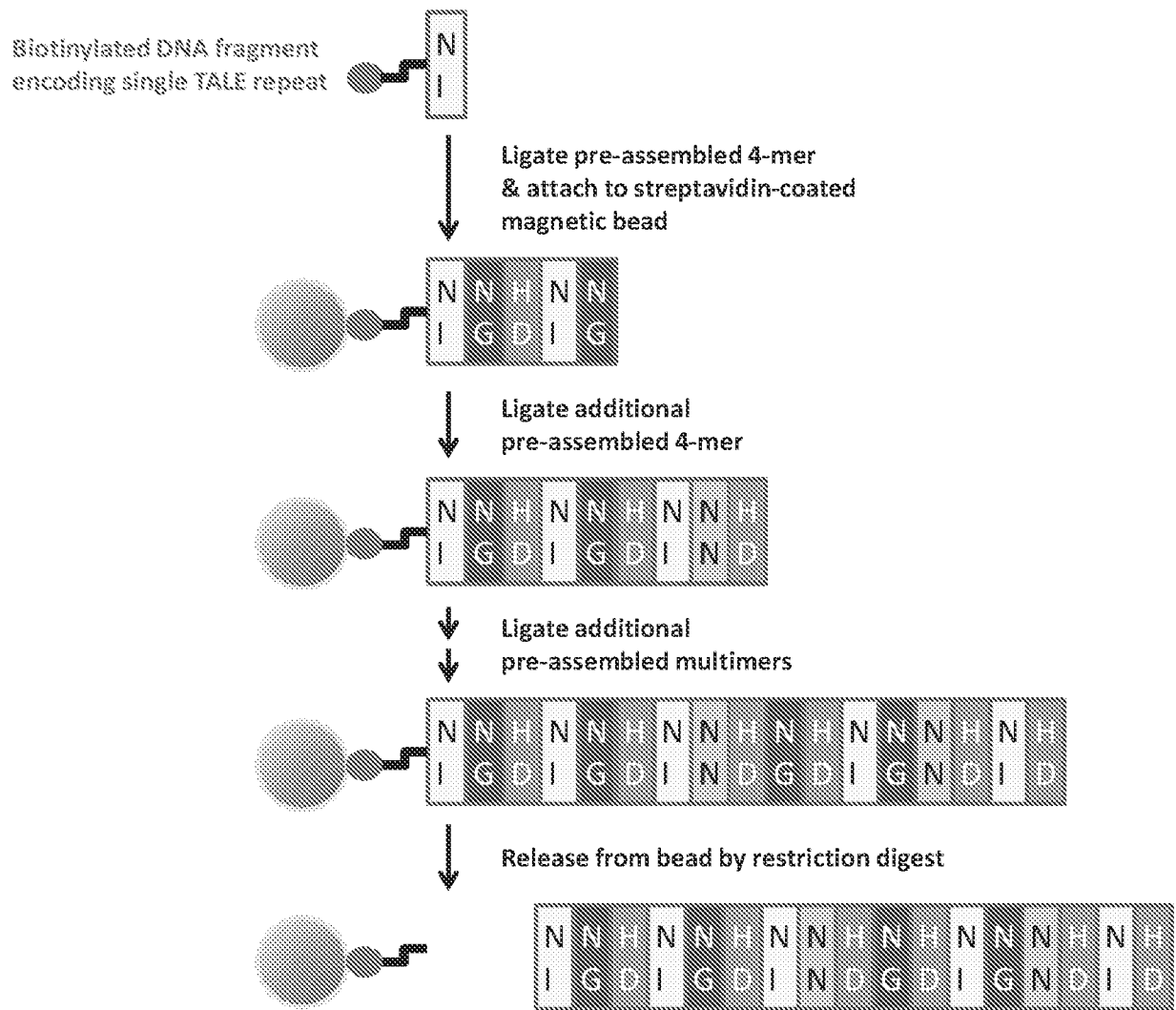


FIG. 1

Archive of 376 pre-assembled TALE repeat units

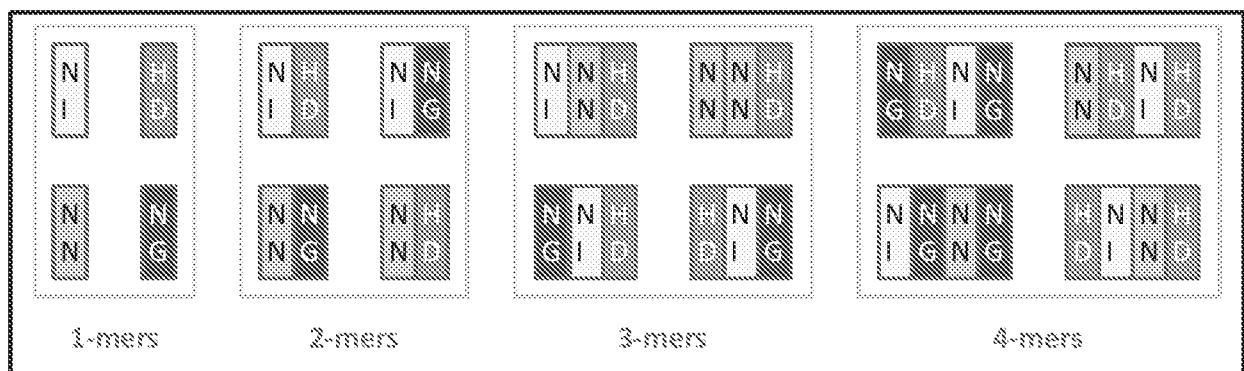


FIG. 2

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGT
CTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGGG
CTGGCTTAACCTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGC
ACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCCATTCCGCCATTCAGGCTGCGCAACTGTTGGGAA
GGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATT
AAGTTGGGTAAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCG
GTACCTCGCGAATGCATCTAGATATCGGATCCCGGGCCCCGTCGACTGCAGAGGCCTGCATGCAAGCTT
GGCGTAATCATGGTCATAGCTGTTTCCCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAGCCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGTCACTGCCCGCTTTCAGTCGGGAAACCTGTGCTGCCAGCTGCATTAATGAATCGGCCAACGCGC
GGGGAGAGGCGGTTTTGCGTATTTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCG
TTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT
GGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGC
GAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTT
CCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGATAGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTACAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGTAAGACACGAC
TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGA
AGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
TTCTACGGGTCTGACGCTCAGTGGAACGAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA
AAAGGATCTTCACCTAGATCCTTTTAAATTTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTGCG
TTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCC
CCAGTGCTGCAATGATACCGCGAGAGCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCA
GCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTG
CCGGGAAGCTAGAGTAAGTAGTTCCGCAAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGCTTCCCAACGATCAAGGCGAGTT
ACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCG
TAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCG
AGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCAT
CATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGT
AACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAA
ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT
CCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA
TTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAA
ACCATATTATCATGACATTAACTATAAAAAATAGGCGTATCACGAGGCCCTTTTCGT
(SEQ ID NO:1)

FIG. 3

FIG 4A

Type	SEQUENCE	SEQ ID NO:
α/ϵ	LTP D QVVVAIASNIGGKQALETVQRLLPVLCQ D HG	2
β	LTP E QVVVAIASNIGGKQALETVQRLLPVLCQ A HG	3
γ	LTP D QVVVAIASNIGGKQALETVQRLLPVLCQ A HG	4
δ	LTP A QVVVAIASNIGGKQALETVQRLLPVLCQ D HG	5

FIG 4B

Type	SEQUENCE	SEQ ID NO:
α/ϵ	CTGACCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGG GGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCG GTCCTTTGTCAAGACCACGGC	6
Type β :	CTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGG TGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCAG TTCTCTGTCAAGCCCACGGG	7
Type γ :	CTGACTCCCATCAAGTTGTAGCGATTGCGTCGAACATTGGA GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCCT GTTGTGTCAAGCCCACGGT	8
Type δ :	TTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCG GTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACT GTGCCAGGATCATGGA	9

GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTA
TCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGAC
CGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTG
ACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTA
CATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCC
CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCACTTGGCAGTACAT
CAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCAGTACA
TGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGATAGCGGTTTGACTCACGGGGATTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTTGACGCAAATGGGCGGTAGGCGT
GTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAAT
ACGACTCACTATAGGGAGACCCAAGCTGGCTAGCaccATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGA
CATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTACCCGCGGGGTACCT
ATGGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTGAGGAGCACCGTCGC
GCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCTGCGGCGCT
TGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCGGAAGCCACGCACGAGGCAATTGTAGGGGTG
GTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCA
GCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGACGCCTGGCG
CAATGCGCTCACCGGGGCCCCCTTGAACAGAGACGATTAATGCGTCTCGCTGACACCCGAACAGGTGGTCGCCATTGC
TXXXXXXXXXXGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTT
AACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCA
TGCTCCCGCATTGATCAAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCACTAGT
CAAAAGTGAAGTGGAGGAGAAGAAATCTGAAGTTCGTCATAAATTGAAATATGTGCCTCATGAATATATTGAATTAAT
TGAAATTGCCAGAAATTCCTCAGGATAGAATTCTTGAATGAAGGTAATGGAATTTTTATGAAAGTTTATGGATAT
AGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTGCGATCTCTATTGATTACGGTGTG
ATCGTGGATACTAAAGCTTATAGCGGAGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTCGAA
GAAAATCAAACACGAAACAAACATATCAACCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAATTTAAGT
TTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGA
GCTGTTCTTAGTGTAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGA
CGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCG
ATTCTACGCGTACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCACTGTCTTCTCTAATAAAAT
GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTG
GGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTG
ACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCTTCTCGCCACGTTGCGCGGCTTCCC
CGTCAAGCTCTAAATCGGGGCATCCCTTAGGGTTCCGATTTAGTGCTTACGGCACCTCGACCCCAAAAACCTTGATT
AGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTGTACGTTGGAGTCCACGTTCTTTAA
TAGTGGACTCTTGTTCAAAACCTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTGATTTATAAGGGATTTTGGGGA
TTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTA
GGGTGTGGAAGTCCCCAGGCTCCCCAGGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGT
GTGGAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTC

(SEQ ID NO:10)

FIG. 5A

FIG. 5B

AGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGCCCATGGC
TGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTT
TTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGATATCCATTTTCGGATCTGATCAGCACGTGTTGACA
ATTAATCATCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGCCTTTGTCTC
AAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCAACAGCATCCCCATCTCTGAAGACTACAGCGTCGCC
AGCGCAGCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATATCATTTTACTGGGGGACCTTGTGCAGAA
CTCGTGGTGCTGGGCACTGCTGCTGCTGCGGCAGCTGGCAACCTGACTTGTATCGTCGCGATCGGAAATGAGAACA
GGGGCATCTTGAGCCCCTGCGGACGGTGTCGACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAA
GGACAGTGATGGACAGCCGACGGCAGTTGGGATTCTGTAATTGCTGCCCTCTGGTTATGTGTGGGAGGGGCTAAGCA
CTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTTCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCT
TCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCACCCC
AACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCAC
TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTT
GGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAG
CATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCA
GTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGC
TCTTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTGTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGG
AACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTC
AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTC
CTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCCCGACC
GCTGCGCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTG
GTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACAC
TAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTCATCCATAGTTG
CCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGA
GACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTGGTCCTG
CAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGC
GCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGGCTTCATTACGCTCCGGTTCCC
AACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCA
GAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTATGCCATCCGTAA
GATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAGTGCTCATATTGGAAAACGTTCTTCGGGG
CGAAAACCTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCCACTCGTGCACCCCACTGATCTTCAGCA
TCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGA
CACGGAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGG
ATACATATTTGAATGATTTAGAAAAATAAACAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGT
C (SEQ ID NO:11)

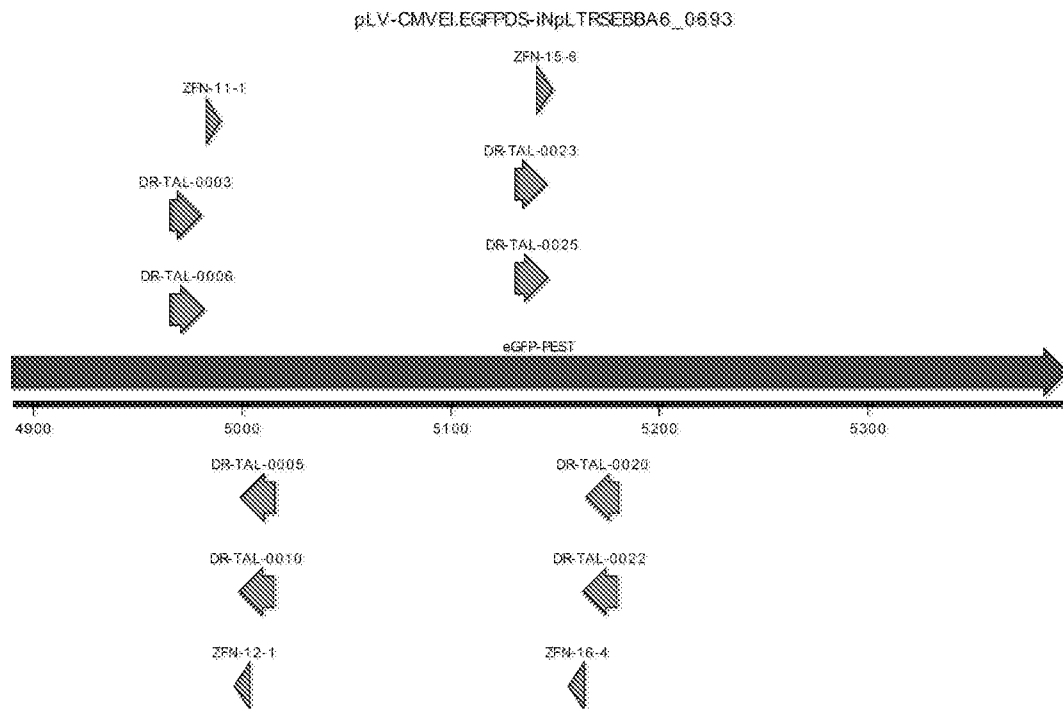


FIG. 6

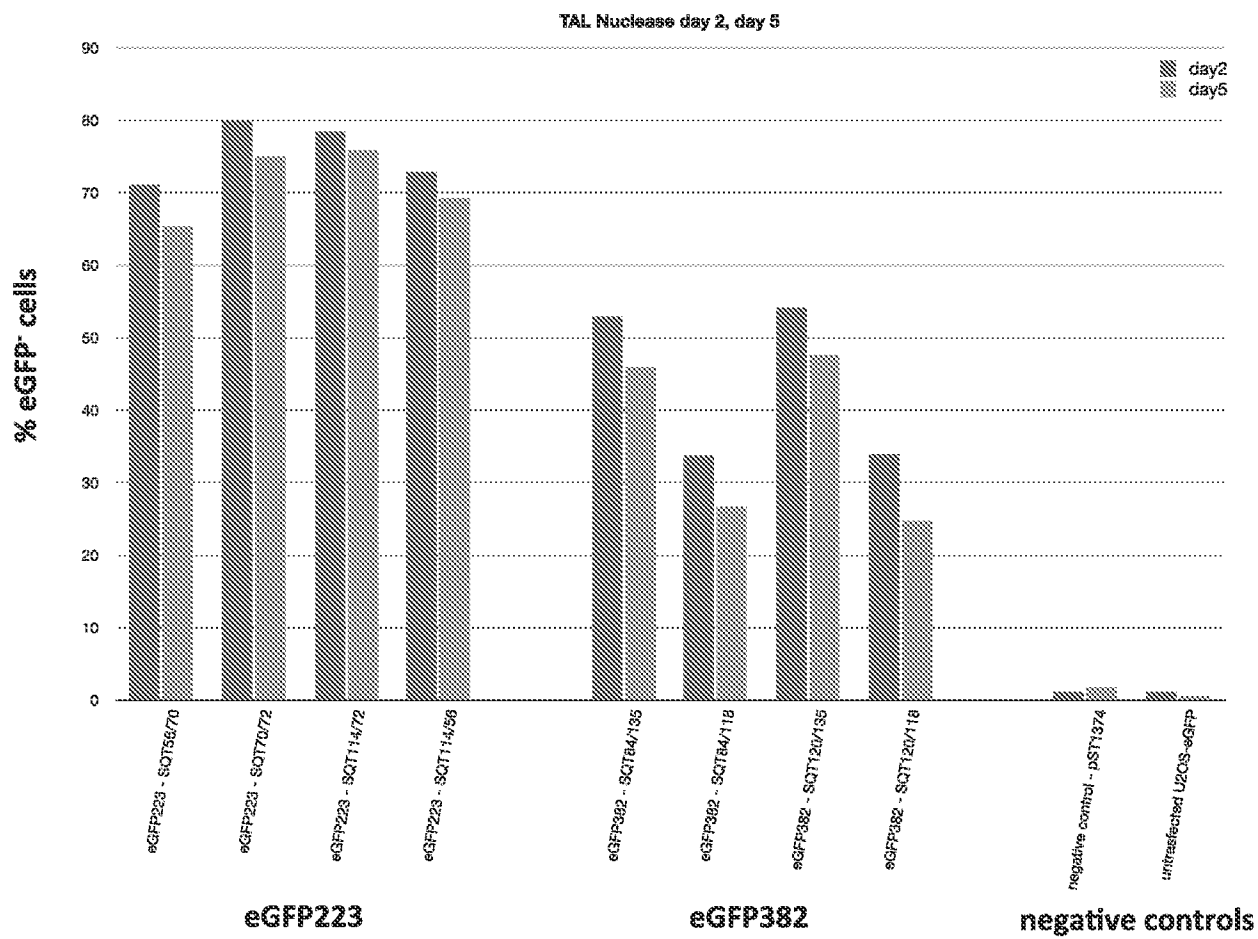
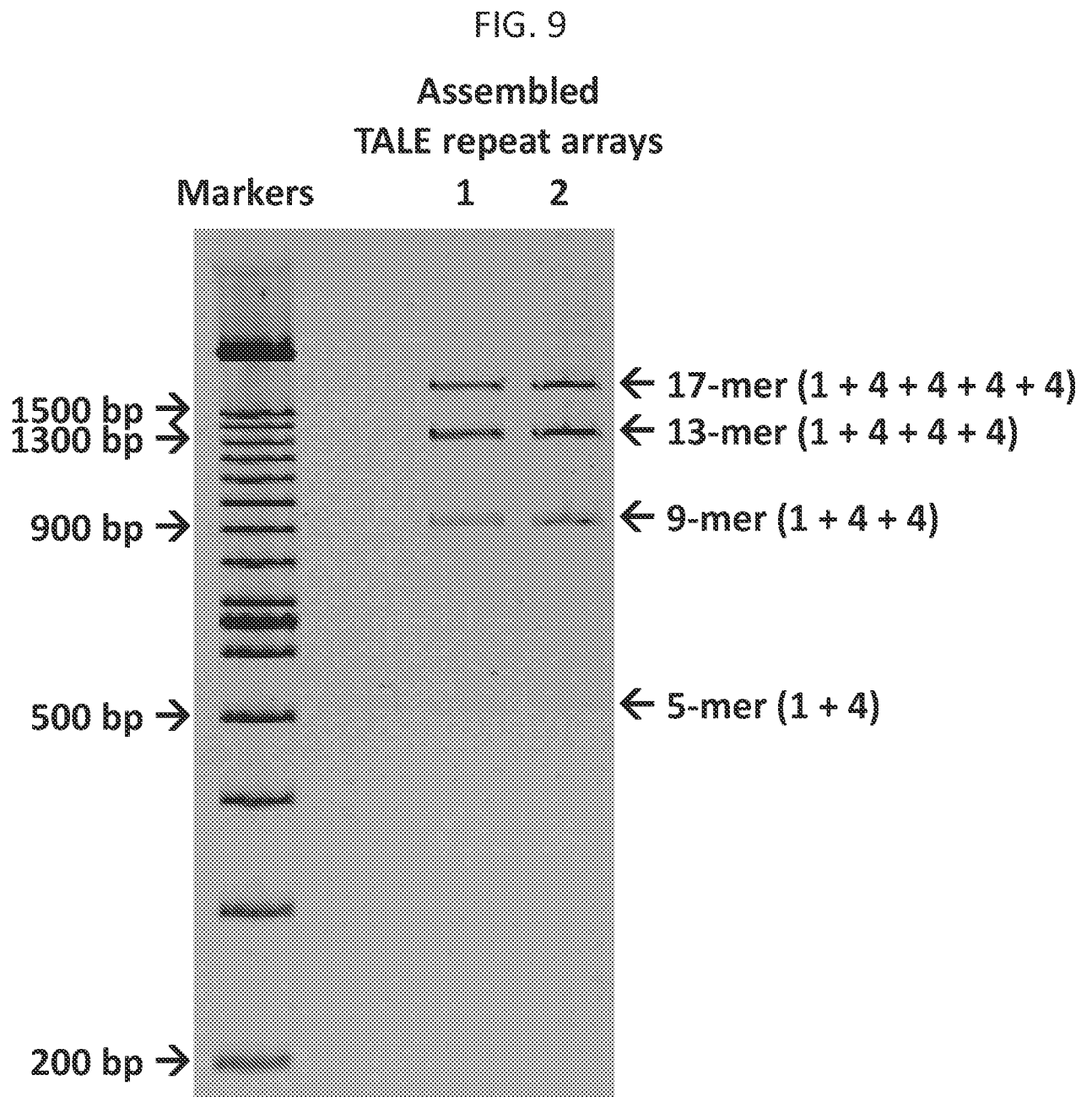


FIG. 7

[illegible]

FIG. 8



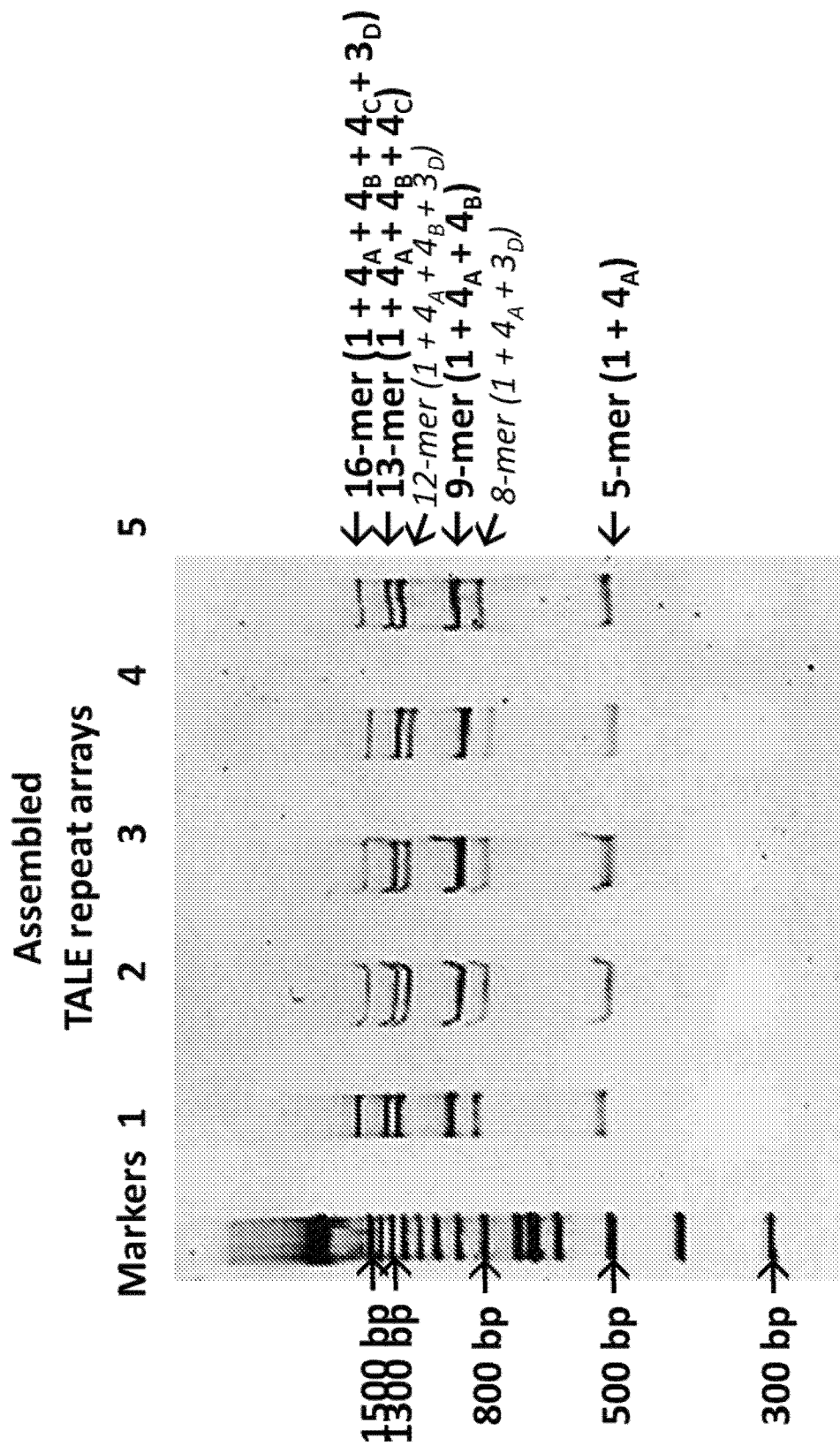


FIG. 10

FIG. 11A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTGAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCTCCGCTCCAGCTCGACA
CCGGGCGAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCACTGCACGCCTGGCGCAATGC
GCTACCGGGGGCCCCCTTGAACCTGACCCAGACAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAA
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTC
GCAATCGCGTCAAACGAGGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGTCAAACGAGGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCAC
GACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGT
CACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTCTGACACCCGAACAGGTGGTCGCCAT
TGCTTCCACGACGGAGGACGGCCAGCCTTGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCT
GCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGT
CTGCCTCATGCTCCCGCATTGATCAAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGAT
CC

(SEQ ID NO:25)

FIG. 11B

MDYKDHGDYKDHIDYKDDDDKMAPKKRKVGIHRGVPMVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
GKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPEQVVAI
ASHDGGRPALLESIVAQLSRPDPALALTNHDLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS

(SEQ ID NO:26)

FIG. 12A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTGAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTGCGCTTTACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGCACGAGGCAATTGTAGGGGTGCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGACGCCTGGCGCAATGC
GCTACCGGGGGCCCCCTTGAACCTGACCCAGACAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTC
GCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCAC
GACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGT
CACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCAT
CGCCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATG
GACTGACACCCGAACAGGTGGTCGCCATTGCTTCTAATGGGGGAGGACGGCCAGCCTTGAGTCCATCGTAGCCC
AATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGAC
GACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCC
CCGAGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:27)

FIG. 13A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCCAAGAAGAAGAGGAAGGTGGGCATTACCCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCCAAGCCACGCACGAGGCAATTGTAGGGGTGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCACTGCACGCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTGTGTCAAGCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAATATTGGCGGTAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTC
GCAATCGCGAACAATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCAGTTCTCTGTCAAGCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCACGA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTG
TGTCAGCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCCAGACCAGGTAGTCGCAATCGCGAAC
AATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCA
GTTCTCTGTCAAGCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCG
CCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTCGCCATTGCTTCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCAGGCCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCATCGAGTCGCGGGATCC
(SEQ ID NO:29)

FIG. 13B

MDYKDHGDYKDHIDYKDDDDKMAPKKRKVGIHRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDGTGQLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
DHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQV
VAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGK
QALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIAN
NNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRPALESIVAQLSRPDPALAALTNHDLVALACLGGRPALDA
VKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:30)

FIG. 14A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGCACGAGGCAATTGTAGGGGTGCGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCGAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGACAGCCTGGCGCAATGC
GCTACCCGGGGCCCCCTTGAACCTGACCCAGACAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTGCCTATCGCCAACAACAACGGCGGTAAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTC
GCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTGCCTATCGCCTCGAATGGCGGCGGTAAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTTG
TGTCAGCCCACGGTTTGACGCCTGCACAAGTGGTGCCTATCGCCAACAACAACGGCGGTAAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTGCCTATCG
CCAGCCATGATGGCGGTAAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTGCCTATTGCTTCTAATGGGGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCAGGCCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCCATCGAGTCGCGGGATCC

(SEQ ID NO:31)

FIG. 15A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTACAGCACCCCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCCCAAGCCACGCACGAGGCAATTGTAGGGGTGCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGC
GCTCACCGGGGGCCCCCTTGAACCTGACCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGC
CCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATT
GCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGG
CTGACTCCCGATCAAGTTGTAGCGATTGCGTCAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTC
CTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGC
AGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCG
CAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACC
ACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCTCCAATATTGGCGGTAAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTG
TGTCAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTCTGACACCCGAACAGGTGGTCGCCATTG
CTTCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCCGCGTTGGCTG
CGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAGGGTC
TGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCGAGAGAACTTCCATCGAGTCGCGGGATC
C

(SEQ ID NO: 33)

FIG. 15B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIA
SHDGGRPALLESIVAQLSRPDPALAAALTNDHLVAIACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVAGS
(SEQ ID NO:34)

FIG. 16A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTGAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGCACGAGGCAATTGTAGGGGTGCGGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGACGCCTGGCGCAATGC
GCTCACCGGGGCCCCCTTGAACCTGACCCAGACCAGGTAGTCGCAATCGCGTCACATGACGGGGGAAAGCAAGC
CCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATT
GCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGG
CTGACTCCCGATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTC
CTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGC
AGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCG
CAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACC
ACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTTCA
GACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGAATAACAATGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCGAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTG
TGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAACGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCA
CATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCACGGCCTTACA
CCGGAGCAAGTCGTGGCCATTGCAAGCAACATCGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCA
GTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCA
TTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCG
CCAGCCATGATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGAC
TGACACCCGAACAGGTGGTCGCCATTGCTAATAATAACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAAT
TGTCAGGCCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGAC
CCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCCG
AGAGAACTTCCCATCGAGTCGCGGGATCC
(SEQ ID NO:35)

FIG. 16B

MDYKDHDGDYKDHDIYKDDDDKMAPKKRKVGHRGVPMVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAGR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDH
GLTPEQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
SHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGRPALSIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:36)

FIG. 17A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGG
CGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGCTTTCACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGCACGAGGCAATTGTAGGGGTGCGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCTCCGCTCCAGCTCGACA
CCGGGCGAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGACGCCTGGCGCAATGC
GCTACCGGGGGCCCCCTTGAACCTGACCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGG
GCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCT
CCTTCCCGTGTGTGTCAAGCCACGGTTTGACGCCTGCACAAGTGGTGCCTATCGCCAACAACAACGGCGGTAAAG
CAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTC
GCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGAC
CACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCAG
AGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCGCATGACGGAG
GGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGT
GGTGCCTATCGCCAGCCATGATGGCGGTAAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCAACATTGGGGGAAAGCAAGCCCTGGAAACCG
TGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAATAA
CGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGAT
CAAGTTGTAGCGATTGCGAATAACAATGGAGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTG
TGTCAGCCCACGGTTTGACGCCTGCACAAGTGGTGCCTATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAA
ACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACCAGGTAGTCGCAATCGCGTCA
AACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTAC
ACCGGAGCAAGTCGTGGCCATTGCAAATAATAACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCC
AGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGAAACAAGC
ATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTGCCTATC
GCCTCGAATGGCGGCGGTAAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGG
ACTGACACCCGAACAGGTGGTGCCTATTGCTAATAATAACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCA
ATTGTCCAGGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACG
ACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCCC
GAGAGAACTTCCCATCGAGTCGCGGGATCC (SEQ ID NO:37)

FIG. 17B

MDYKDHGDYKDHDIYKDDDDKMAPKKRKVGIHRGVPMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDGTQLLKIAKR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIANNNGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQR
LLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
SNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIANNNGRPALESIVAQLSRPDPAALAALNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:38)

FIG. 18A

TCTAGAGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGAC
GATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCCGCGGGGTACCTATGGTGGACTTGAGGAC
ACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTGAGGAGCACCGTCGCGCAACACCCAGAGG
CGCTTGTTGGGGCATGGCTTCACTCATGCGCATATTGTGCGCTTTACAGCACCTGCGGCGCTTGGGACGGTGGC
TGTCAAATACCAAGATATGATTGCGGCCCTGCCGAAGCCACGACGAGGCAATTGTAGGGGTGCGTAAACAGTG
GTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGCCTCCGCTCCAGCTCGACA
CCGGGCGAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAAGTGCACGCCTGGCGCAATGC
GCTACCCGGGGCCCCCTTGAACCTGACCCAGACAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAG
CCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCA
TTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACG
GGCTGACTCCCAGTCAAGTTGTAGCGATTGCGAATAACAATGGAGGGGAAACAAGCATTGGAGACTGTCCAACGGC
TCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCCATGATGGCGGTAA
GCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGT
CGCAATCGCGTCACATGACGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGA
CCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCATCCACGACGGTGGCAAACAGGCTCTTGAGACGGTTCA
GAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCAGTCAAGTTGTAGCGATTGCGTCGCATGACGGA
GGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAA
GTGGTCGCCATCGCCTCCAATATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGAACAATAATGGGGGAAAGCAAGCCCTGGAAAC
CGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTACACCGGAGCAAGTCGTGGCCATTGCAAATAAT
AACGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCCAGTTCTCTGTCAAGCCCACGGGCTGACTCCCG
ATCAAGTTGTAGCGATTGCGTCGAACATTGGAGGGGAAACAAGCATTGGAGACTGTCCAACGGCTCCTTCCCGTGT
TGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGCGGCGGTAAGCAGGCGCTGG
AAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATGGACTGACCCAGACAGGTAGTCGCAATCGCGA
ACAATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCTTTGTCAAGACCACGGCCTTA
CACCGGAGCAAGTCGTGGCCATTGCAAGCAATGGGGGTGGCAAACAGGCTCTTGAGACGGTTCAGAGACTTCTCC
CAGTTCTCTGTCAAGCCCACGGGCTGACTCCCGATCAAGTTGTAGCGATTGCGTCCAACGGTGGAGGGGAAACAAG
CATTGGAGACTGTCCAACGGCTCCTTCCCGTGTGTGTCAAGCCCACGGTTTGACGCCTGCACAAGTGGTCGCCAT
CGCCAACAACAACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGATCATG
GACTGACACCCGAACAGGTGGTCGCCATTGCTTCCACGACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCC
AATTGTCCAGGCCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGAC
GACCCGCGCTCGATGCAGTCAAAAAGGCTGCTCATGCTCCCGCATTGATCAAAAGAACCAACCGGCGGATTCC
CCGAGAGAACTTCCCATCGAGTCGCGGGATCC (SEQ ID NO:39)

FIG. 18B

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHRGVPMVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHG
FTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAGR
GGVTAVEAVHAWRNALTGAPLNLTDPQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASNNGGKQALE
TVQRLLPVLCQAHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGG
KQALETVQRLLPVLCQDHGLTPEQVVAIANNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRL
LPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHG
LTPEQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIA
NNNGGKQALETVQRLLPVLCQDHGLTPEQVVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLPHAPALIKRTNRRIPERTSHRVAGS (SEQ ID NO:40)

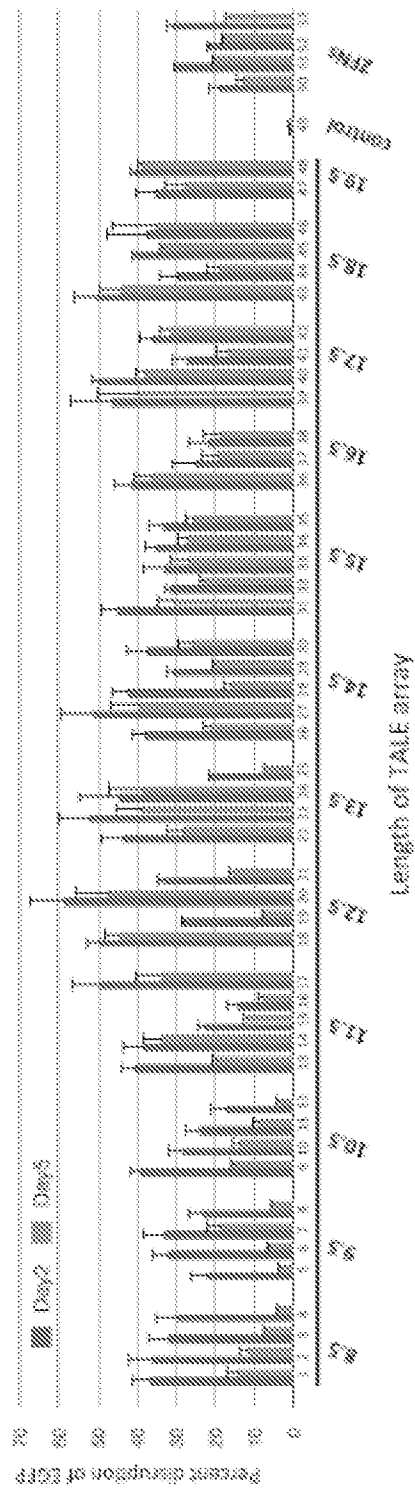


FIG. 19A

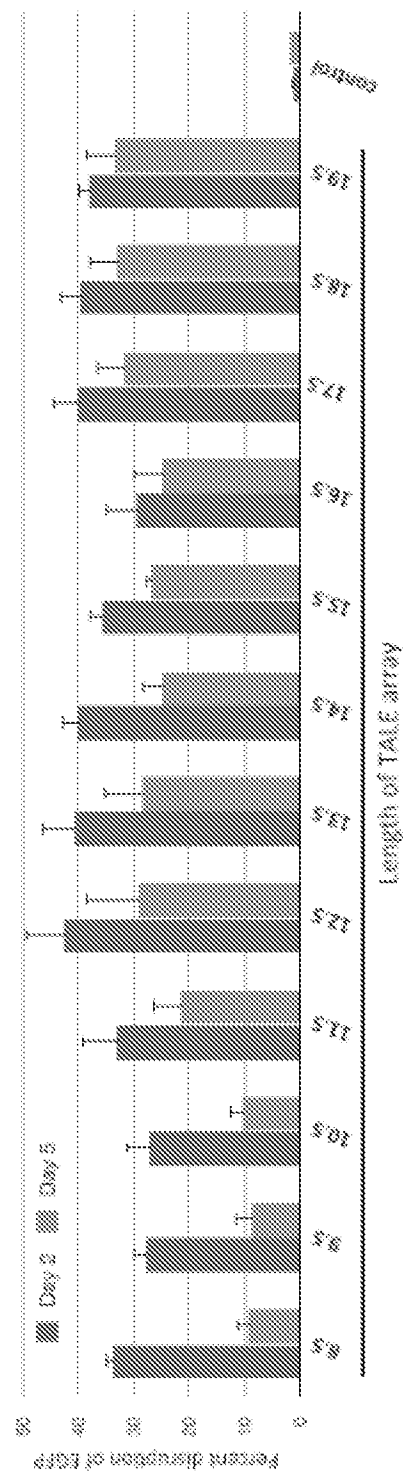


FIG. 19B

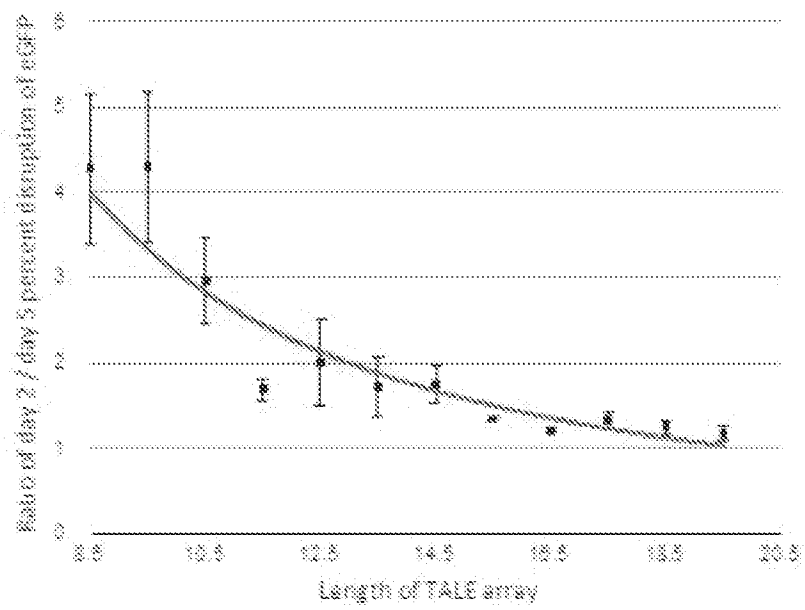


FIG. 20A

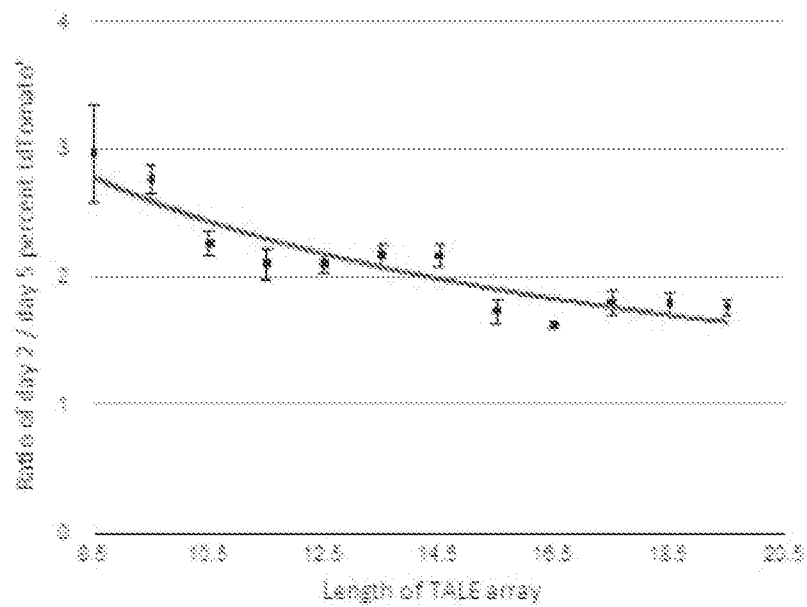


FIG. 20B

SEQUENCE		Δ SIZE	SEQ ID NO:
AXIN2			
TAL2200/TAL2201			
Mutations in 2 of 83 sequences ≈ 2.4%			
TTCCAGACTCAGTGGGAGAGCTCCCTCACC	TTGAGTAGCGCTATGTTGGTGACTTGGCTCCCGACCCAGCAGCAGCTTCCGTGA	WT	301.
TTCCAGACTCAGTGGGAGAGCTCCCTCACC	TTGAGTAGCGCTATGTTGGTGACTTGGCTCCCGACCCAGCAGCAGCTTCCGTGA	Δ8	302.
TTCCAGACTCAGTGGGAGAGCTCCCTCACC	TTGAGTAGCGCTATGTTGGTGACTTGGCTCCCGACCCAGCAGCAGCTTCCGTGA	Δ31	303.
BRCA1			
TAL2384/TAL2385			
Mutations in 7 of 14 sequences ≈ 50.0%			
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	WT	304.
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	Δ17	305.
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	Δ23	306.
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	Δ27	307.
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	Δ38	308.
GGCGTGGGAGAGTGGATTTCGGAAGCTGACAGAT	GGGTAGGGGGGTAGGGCGGAACCTGAGAGGCGTAAGCGTTGTG	Δ39	309.
CHD7			
TAL2238/TAL2239			
Mutations in 4 of 81 sequences ≈ 4.9%			
CCTGAGCTCTGCTTGGAGGAGCCGCTGTGTTGGAAGAAGATG	CAGATCCAGGAATGATGAGTCTTTTGGCCGAGGATGGGAATATTT	WT	312.
CCTGAGCTCTGCTTGGAGGAGCCGCTGTGTTGGAAGAAGATG	CAGATCCAGGAATGATGAGTCTTTTGGCCGAGGATGGGAATATTT	Δ9 (Δ11 +2)	313.
CCTGAGCTCTGCTTGGAGGAGCCGCTGTGTTGGAAGAAGATG	CAGATCCAGGAATGATGAGTCTTTTGGCCGAGGATGGGAATATTT	Δ10	314.
CCTGAGCTCTGCTTGGAGGAGCCGCTGTGTTGGAAGAAGATG	CAGATCCAGGAATGATGAGTCTTTTGGCCGAGGATGGGAATATTT	Δ28	315.
TAAAGAAATATGCAATGACAT	AAATCCAGTAAATCCTAIG	Δ199	316.

FIG. 21A

FIG. 21B

SEQUENCE		Δ SIZE	SEQ ID NO:
NCOR2			
TAL2284/TAL2285			
Mutations in 7 of 88 ≈ 8.0%			
TCCACACAGCCTGTGGCACAGACGTGGAGGCCCACTGAGCCCGCTACCCGCCCCACAGCCTTTCTACCCAGTGCAGATCGCCCGGA	WT		337.
TCCACACAGCCTGTGGCACAGACGTGGAGGCCCACTGAGCCCGCTA-CGGCCCCACAGCCTTTCTACCCAGTGCAGATCGCCCGGA	Δ1		338.
TCCACACAGCCTGTGGCACAGACGTGGAGGCCCACTGAGCCCGC-----CCCACAGCCTTTCTACCCAGTGCAGATCGCCCGGA	Δ7		339.
TCCACACAGCCTGTGGCACAGACGTGGAGGCCCACTGACC-----CCCACAGCCTTTCTACCCAGTGCAGATCGCCCGGA	Δ12		340.
TCCACAC-----AGATCGCCCGGA	Δ69		341.
CCCCACCACC-----/-----TTTCCTACCCAGTGCAGATCGCCCGGA	Δ70		342.
GCTTATTGGGC-----/-----CTTTCTCCTGGAGCCAGGCT	Δ213		343.
TCCACACAGCCTGTGGCACAGACGTGGAGGCCCACTGAGCCCGCTACAGCCGCCCCACAGCCTTTCTACCCAGTGCAGATCGCC	+3		344.
JAK2			
TAL2406/TAL2407			
Mutations in 13 of 21 sequences ≈ 61.9%			
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATGCGAATGGCCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT	WT		345.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATG-----GGGCCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ4		346.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATG-----GGCCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ5		347.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATGGATGG-----TTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ6		348.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCA-----GCCCTGCCTTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ8		349.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGC-----CTTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ15 (2x)		350.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATGG-----AATGGAGGGAACATCCACCTCT	Δ24		351.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGAC-----GATGACAGAAATGGAGGGAACATCCACCTCT	Δ25		352.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTG-----CAGAAATGGAGGGAACATCCACCTCT	Δ26		353.
TTTCTCTTACAGGC-----ATGGAGGGAACATCCACCTCT	Δ53		354.
T-----ACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ54		355.
TTTCTCTTACAGGCAAAATGTTCTGAAAAAGACTCTGCATGG-----aglaadtCTTACGATGACAGAAATGGAGGGAACATCCACCTCT	Δ4 (Δ11 +7)		356.
CGGAGGTTTCTGCAaacacagaaattt-----/-----CGATGACAGAAATGGAGGGAACATCCACCTCT	Δ288 (Δ301 +13)		357.

FIG. 21C

SEQUENCE		Δ SIZE	SEQ ID NO:
MYCN			
TAL2280/TAL2281			
Mutations in 12 of 35 sequences ~ 34.3%			
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	WT		358.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ6 (Δ8 +2)		359.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ8		360.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ10		361.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ11		362.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ12		363.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ18		364.
CGGAGGCGGAGCCCGATGTCGAGCTGCTCCAGTCCACCATGCCGGGCGATGATCTGCAAGAACCCAGACCTCGAGTTTGACTCGCTACA	Δ19		365.
AAGGAAGCAGCCCGGTCCTTAA / /	Δ100		366.
ACTGTTGAGGTCGGCGCGCCGCCCG / /	Δ122		367.
AAGGAAGCAGCCCGGTCCTTAA / /	Δ141		368.
GAGCGAGCCGATGCCCGGTCGATCTGGGTACGGGCTGCTCCAGCTGGAGGAGGGGGCTCTCCGGGAGCCCTCCTCGCGGGGGGGC	+160 (Δ29 +189) (2x)		369.
CCCTGCCATTCGCCGGAACAGGGGCTCAGGCCCTTTGTAGTGCTCGTATGCTTGGCTGGGAGCATTTTGGAGGCGAGTGCTAGG			
GGCAGAGAGGTCCTGTTTCCCCCAAGTCTGATCTGCAAGAACCCCA			
NBN			
TAL2408/TAL2409			
Mutations in 12 of 20 sequences ~ 60%			
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	WT		370.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ5		371.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ7		372.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ13		373.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ17 (Δ19 and +2)		374.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ19		375.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ28		376.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ28		377.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ30		378.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ35		379.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ39		380.
TGCACGTGCGGCCCCAGCCCTGAGGAGCCGGACCGATGTTGGAAACTGCTGCCGCCCGCGGCGCCGCGCAGGAGGTAAGGGCAGAAGGGAA	Δ45		381.
GCTCCCGGAGGCGCACGTCCCGGAGCCCAT / /	Δ183		382.

FIG. 21D

SEQUENCE		Δ SIZE	SEQ ID NO:
XPC - TAL2350/TAL2351			
Mutations in 10 of 36 ≈ 27.8%			
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGCTCGGAACGCGCGCGCGGGAGCCCGGGGACCGCGAATCGCGAGCCAG	WT		383.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGCTCGGAGG--AACGGCGCGCGGGAGCCCGGGGACCGCGAATCGCGAGCCAG	Δ 2 (Δ6 and +4)		384.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGCTCGGAAA-----GGGAGCCCGGGGACCGCGAATCGCGAGCCAG	Δ12		385.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGGC-----CGCGGGAGCCCGGGGACCGCGAATCGCGAGCCAG	Δ15		386.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGGCTCG-----GGGAGCCCGGGGACCGCGAATCGCGAGCCAG	Δ17 (x2)		387.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGGC-----GGGAGCCCGGGGACCGCGAATCGCGAGCCAG	Δ19		388.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGTT-----GCGAATCGCGAGCCAG	Δ35 (Δ37 and +2)		389.
TCGCGAAGTGAATTTGCCAGACAAAGCAACATGGCTCG-----GCGAGCCAG	Δ39		390.
TCGCG-----CAGCCAG	Δ76		391.
TCGCGAAGTGG-----/ /-----AATCAAGGCCAAGCAAG	Δ78		392.
BRCC2 - TAL2306/TAL2307			
Mutations in 19 of 20 sequences ≈ 95.0%			
GTGACCCCGCTGCACAGTCGGCGCGCGGCTGGGGTCCGCTCGCTAGCGGCGCGGGGGTCTTGAA	WT		393.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAAGT-----GGGCTGGGGTCCGCTCGCTAGCGGCGCGGGGGTCTTGAA	Δ7		394.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAA-----GCTGGGGTCCGCTCGCTAGCGGCGCGGGGGTCTTGAA	Δ11		395.
GTGACCCCGCTGCACAGTCGGCGCGCGGCG-----TCGCTAGCGGCGCGGGGGTCTTGAA	Δ31		396.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAAGT-----GGGGTCTTGAA	Δ39		397.
GGCCTCTCGTGAATATTATG-----/ /-----ATGGGTCTATCGGTGGCGCG	Δ99		398.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAAG-----/ /-----GGTGAAGGGTGGAC	Δ100		399.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAAGT-----/ /-----CGTCGGGTGGACCGGCGCAGCCG	Δ99 (Δ104 + 5)		400.
GTGACCCCGCTGCACAGTCGGCGCGCGGCGCATGAAGT-----/ /-----GGCAGCCGAGGAGCAGCAGT	Δ118		401.
ATGGTCTCGTAATATAGTGGAGCGAGCCCTCGAGG-----/ /-----GGGGTCTTGAAGATG	Δ122		402.
GTGACCCCGCTG-----/ /-----CTGAGGACCTGAGGGTTACC	Δ200		403.
GGCCTCTCGTGAATATTCA-----/ /-----GCTCCTGAGGACCTGAGGGTTA	Δ221		404.
GTGACCCCGCTGCACAGTCGGC-----/ /-----CGTCACCCCTCTCTGGGCT	Δ222 (2x)		405.
GTGACCCCGCTGCACAGTCGGC-----/ /-----GCTCGAGCAGCCGGG	Δ238		406.
GTGACCCCGCTGCACAGTCGGCGCGC-----/ /-----GACCGGCACTGTGAGG	Δ242		407.
GTGACCCCGCTGCACAGTCGGC-----/ /-----CGGGCACTGTGAGG	Δ249		408.
GTGAC-----/ /-----CCAGCAGCCCTG	Δ313		409.
GTGACCC-----/ /-----CCTGCCGCA	Δ319		410.
GGCCTCTCGTGAATATTATGAG-----/ /-----CCGCGAGGCTCAACCTGGA	Δ340		411.

FIG. 21E

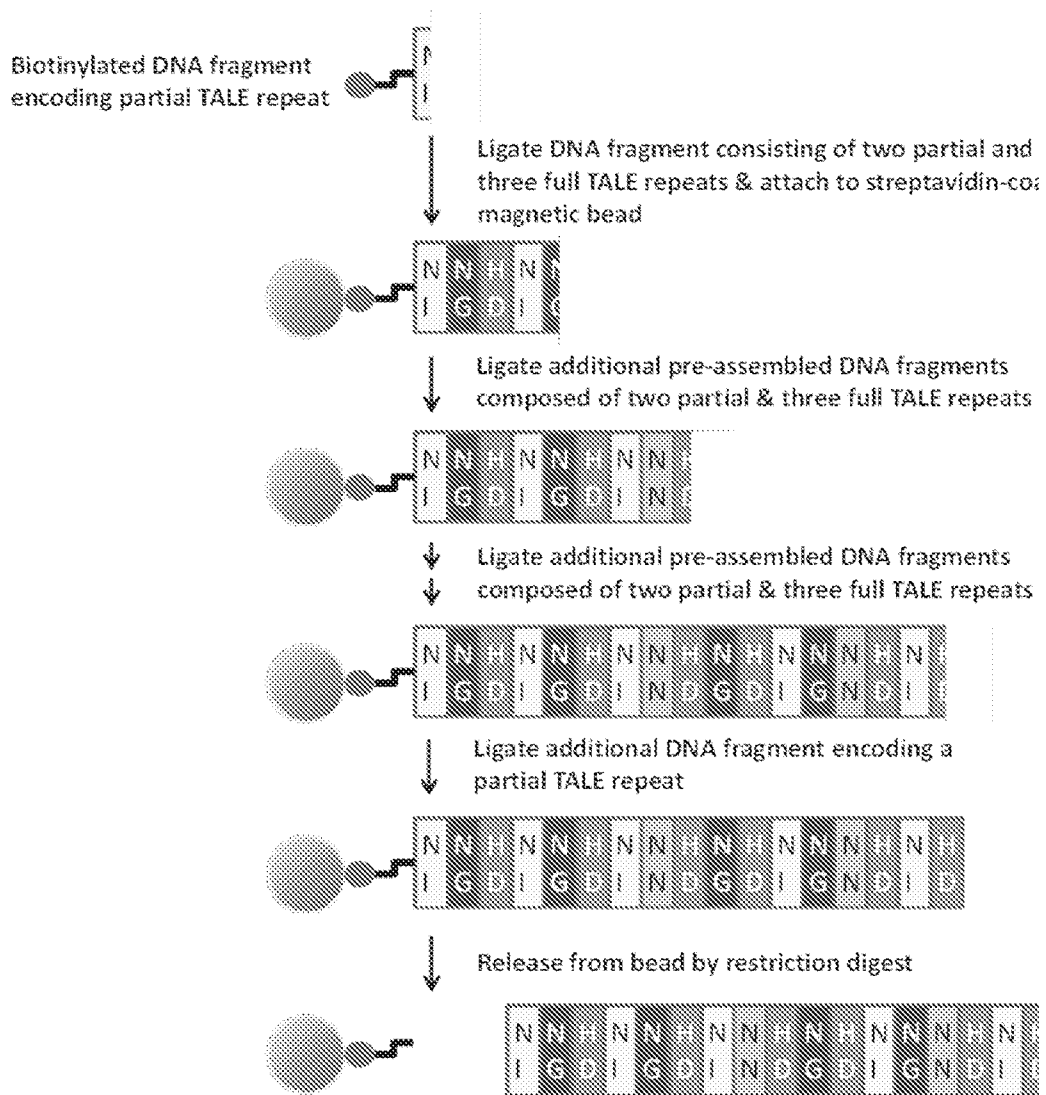


FIG. 22

29539_0036W01_SL2953936W.txt
SEQUENCE LISTING

<110> THE GENERAL HOSPITAL CORPORATION

<120> METHODS OF TRANSCRIPTION ACTIVATOR LIKE EFFECTOR ASSEMBLY

<130> 29539-0036W01

<140>

<141>

<150> 61/610,212

<151> 2012-03-13

<150> 61/601,409

<151> 2012-02-21

<150> 61/508,366

<151> 2011-07-15

<160> 439

<170> PatentIn version 3.5

<210> 1

<211> 2710

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 1

tcgcgcggtt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg	120
ttggcggggtg tcgggggttg cttactatg cggcatcaga gcagattgta ctgagagtgc	180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc	240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggccc tcttcgctat	300
tacgccagct ggcgaaaggg ggatgtgtcg caaggcgatt aagttgggta acgccagggg	360
tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggg acctcgcgaa	420
tgcatctaga tatcggtacc cgggcccgtc gactgcagag gcctgcatgc aagcttggcg	480
taatcatggg catagctgtt tcctgtgtga aattgttata cgctcacaat tccacacaac	540
atacgagccg gaagcataaa gtgtaaagcc tgggggtgcct aatgagttag ctaactcaca	600
ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat	660
taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggagctc ttccgcttcc	720
tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc gagcgggtatc agctcactca	780
aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca	840
aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggt tttccatagg	900
ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagagggt gcgaaaccgc	960
acaggactat aaagatacca ggcgtttccc cctggaagct ccctcggtcg ctctcctgtt	1020

29539_0036W01_SL2953936W.txt

ccgaccctgc cgcttacccg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt	1080
tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc	1140
tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt	1200
gagtccaacc cggtaagaca cgacttatcg ccaactggcag cagccactgg taacaggatt	1260
agcagagcga ggtatgtagg cgggtgctaca gagttcttga agtggtggcc taactacggc	1320
tacactagaa gaacagtatt tggtatctgc gctctgtcga agccagttac cttcggaaaa	1380
agagttggta gctcttgatc cggcaaacaa accaccgctg gtagcgggtg tttttttgtt	1440
tgcaagcagc agattacgcg cagaaaaaaaa ggatctcaag aagatccttt gatcttttct	1500
acgggggtctg acgctcagtg gaacgaaaac tcacgttaag ggatttttgt catgagatta	1560
tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa	1620
agtatatatg agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc	1680
tcagcgatct gtctatttcg ttcattcata gttgcctgac tccccgtcgt gtagataact	1740
acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcg agagccacgc	1800
tcaccggctc cagatttatc agcaataaac cagccagccg gaagggccga gcgcagaagt	1860
ggctctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta	1920
agtagttcgc cagttaatag tttgcgcaac gttgttgcca ttgctacagg catcgtggtg	1980
tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt	2040
acatgatccc ccatggttg caaaaaagcg gttagctcct tcggtcctcc gatcgttgctc	2100
agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt	2160
actgtcatgc catccgtaag atgcttttct gtgactgggt agtactcaac caagtcattc	2220
tgagaatagt gtatgcggcg accgagttgc tcttgcccg cgtcaatacg ggataatacc	2280
gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa	2340
ctctcaagga tcttacgcgt gttgagatcc agttcgatgt aaccactcg tgcacccaac	2400
tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa	2460
aatgccgcaa aaaagggaat aaggcgaca cggaatgtt gaatactcat actcttcctt	2520
tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa	2580
tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccacct	2640
gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg	2700
ccctttcgtc	2710

<210> 2

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 2
 Leu Thr Pro Asp 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> 3
 <211> 34
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 3
 Leu Thr Pro Glu 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 Ala
 20 25 30

His Gly

<210> 4
 <211> 34
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 4
 Leu Thr Pro Asp 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 Ala
 20 25 30

His Gly

<210> 5
 <211> 34
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 5
 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> 6
<211> 102
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

```
<400> 6
ctgacccag accaggtagt cgcaatcgcg tcgaacattg ggggaaagca agccctggaa      60
accgtgcaaa ggttgttgcc ggtcctttgt caagaccacg gc                        102
```

<210> 7
<211> 102
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 7
cttacaccgg agcaagtcgt ggccattgca agcaacatcg gtggcaaaca ggctcttgag 60
acggttcaga gacttctccc agttctctgt caagcccacg gg 102

```
<210> 8
<211> 102
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

```
<400> 8
ctgactcccg atcaagttgt agcgattgcg tcgaacattg gagggaaaca agcattggag      60
actgtccaac ggctccttcc cgtgtttgtgt caagcccacg gt                        102
```

```
<210> 9
<211> 102
<212> DNA
<213> Artificial Sequence
```

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 9
ttgacgcctg cacaagtggg cggcatcgcc tccaatattg gcggtaa gca ggcgctggaa 60

acagtacagc gcctgctgcc tgtactgtgc caggatcatg ga

102

<210> 10

<211> 3269

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<220>

<221> modified_base

<222> (1483)..(1491)

<223> a, c, t, g, unknown or other

<400> 10

gacggatcgg gagatctccc gatcccctat ggtcgactct cagtacaatc tgctctgatg	60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg	120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc	180
ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatagc cgttgacatt	240
gattattgac tagttattaa tagtaatcaa ttacgggggtc attagttcat agcccatata	300
tggagttccg cgttacataa cttacggtaa atggcccggc tggctgaccg cccaacgacc	360
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc	420
attgacgtca atgggtggac tatttacggg aaactgccca cttggcagta catcaagtgt	480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	540
atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca	600
tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg	660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg	780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggctaact agagaaccca	840
ctgcttactg gcttatcgaa attaatacga ctactatag ggagacccaa gctggctagc	900
accatggact acaaagacca tgacggtgat tataaagatc atgacatcga ttacaaggat	960
gacgatgaca agatggcccc caagaagaag aggaaggtgg gcattcaccg cggggtacct	1020
atggtggact tgaggacact cggttattcg caacagcaac aggagaaaat caagcctaag	1080
gtcaggagca ccgtcgcgca acaccacgag gcgcttgtgg ggcattggctt cactcatgcg	1140
catattgtcg cgctttcaca gcaccctgcg gcgcttggga cgggtggctgt caaataccaa	1200
gatatgattg cggccctgcc cgaagccacg cacgaggcaa ttgtaggggt cggtaaacag	1260
tggtcggggag cgcgagcact tgaggcgctg ctgactgtgg cgggtgagct tagggggcct	1320
ccgctccagc tcgacaccgg gcagctgctg aagatcgcg agagaggggg agtaacagcg	1380
gtagaggcag tgcacgcctg gcgcaatgcg ctcaccgggg ccccttgaa cagagacgat	1440
taatgcgtct cgctgacacc cgaacagggt gtcgccattg ctnnnnnnnn nggaggacgg	1500

29539_0036W01_SL2953936W.txt

ccagccttgg agtccatcgt agcccaattg tccaggcccg atcccgcgtt ggctgcgтта	1560
acgaatgacc atctgggtggc gttggcatgt cttgggtggac gaccgcgct cgatgcagtc	1620
aaaaagggtc tgcctcatgc tcccgcattg atcaaaagaa ccaaccggcg gattcccгag	1680
agaacttccc atcgagtcgc gggatcccaa ctagtcaaaa gtgaactgga ggagaagaaa	1740
tctgaacttc gtcataaatt gaaatatgtg cctcatgaat atattgaatt aattgaaatt	1800
gccagaaatt ccactcagga tagaattctt gaaatgaagg таатггаatt ttttatgaaa	1860
gtttatggat atagaggtaa acatttgгgt ggatcaagga aaccggacgg agcaatttat	1920
actgtcggat ctctattga ttacggtgtg atcgtggata ctaaagctta tagcggaggт	1980
tataatctgc caattggcca agcagatgaa atgcaacgat atgtcgaaga aaatcaaaca	2040
cгааасааac atatcaacc таатгаатgg tggaaagtct atccatcttc tgtaacggaa	2100
tttaagtттt татттgtgag tggtcacttt aaaggaaact acaaagctca gcttacacga	2160
ttaaatcata tcactaattg таатггagct gttcttagtg tagaagagct tttaattggt	2220
ggagaaatga ttaaagccgg cacattaacc ttagaggaag tcagacggaa atttaataac	2280
ggcgagataa actтттаagg gcccttcгаа ggtaagccta tccctaacc tctcctcggt	2340
ctcgattcta cgcgtaccgg tcatcatcac catcaccatt gagтттааac ccgctgatca	2400
gcctcgactg tgccttctag ttgccagcca tctgttgттt gcccctccc cgtgccttcc	2460
ttgaccctgg aagggtccac tcccactgtc ctttcctaат aaatgagga aattgcatcg	2520
cattgtctga gtaggtgtca ttctattctg gggggtgggg tggggcagga cagcaagggg	2580
gaggattggg aagacaatag caggcatgct ggggatgcgg tgggctctat ggcttctgag	2640
gcggaaagaa ccagctgggg ctctaggggg tatccccacg cgcctgtag cggcgatta	2700
agcgcggcgg gtgtgggtgt tacgcgcagc gtgaccgcta cacttgccag cgcctagcg	2760
cccgtcctt tcgctttctt cccttccttt ctgccacgt tcgccggctt tcccgtcaa	2820
gctctaaatc ggggcatccc tttagggttc cgatttagtg ctttacggca cctcgacccc	2880
aaaaaacttg attaggtga tggttcacgt agtgggcat cgcctgata gacggtттtt	2940
cgcctттtga cgttgгagtc cacgttcttt aatagtggac tcttgттcca aactggaaca	3000
acactcaacc ctatctcggt ctattctттt gatttataag ggattттggg gatttcggcc	3060
tattggttaa aaaatgagct gatttaacaa aaatttaacg cgaattaatt ctgtggaatg	3120
tgtgtcagtt aggggtgtgga aagtccccag gctccccagg caggcagaag tatgcaaagc	3180
atgcatctca attagtcagc aaccaggtgt ggaaagtccc caggctcccc agcaggcaga	3240
agtatgcaaa gcatgcatct caattagtc	3269

<210> 11

<211> 3178

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 11

agcaaccata gtcccccccc taactccgcc catccccgcc ctaactccgc ccagttccgc	60
ccattctccg ccccatggct gactaatttt ttttatttat gcagaggccg aggccgcctc	120
tgcctctgag ctattccaga agtagtgagg aggctttttt ggaggcctag gcttttgcaa	180
aaagctcccg ggagcttgta tatccatttt cggatctgat cagcacgtgt tgacaattaa	240
tcatcggcat agtatatcgg catagtataa tacgacaagg tgaggaaacta aaccatggcc	300
aagcctttgt ctcaagaaga atccaccctc attgaaagag caacggctac aatcaacagc	360
atccccatct ctgaagacta cagcgtcgcc agcgcagctc tctctagcga cggccgcctc	420
ttcactggtg tcaatgtata tcattttact gggggacctt gtgcagaact cgtggtgctg	480
ggcactgctg ctgctgcggc agctggcaac ctgacttgta tcgtcgcgat cggaaatgag	540
aacaggggca tcttgagccc ctgctggacgg tgtcgacagg tgcttctcga tctgcatcct	600
gggatcaaag cgatagtga ggacagtgat ggacagccga cggcagttgg gattcgtgaa	660
ttgctgccct ctgggttatgt gtgggagggc taagcacttc gtggccgagg agcaggactg	720
acacgtgcta cgagatttcg attccaccgc cgccttctat gaaagggttg gcttcggaat	780
cgttttccgg gacgccggct ggatgatcct ccagcgcggg gatctcatgc tggagttctt	840
cgccaccccc aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac	900
aaatttcaca aataaagcat ttttttact gcattctagt tgtggtttgt ccaaactcat	960
caatgtatct tatcatgtct gtataccgtc gacctctagc tagagcttgg cgtaatcatg	1020
gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc	1080
cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca cattaattgc	1140
gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg tgccagctgc attaatgaat	1200
cggccaacgc gcggggagag gcggtttgctg tattgggcgc tcttccgctt cctcgctcac	1260
tgactcgctg cgctcggtcg ttcggctgctg gcgagcggta tcagctcact caaaggcggg	1320
aatacgggta tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca	1380
gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc	1440
ccctgacgag catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact	1500
ataaagatac caggcgtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct	1560
gccgcttacc ggatacctgt ccgcctttct cccttcggga agcgtggcgc tttctcaatg	1620
ctcacgctgt aggtatctca gttcgggtgta ggtcgttcgc tccaagctgg gctgtgtgca	1680
cgaaccccc gttcagccc accgctgcgc cttatccggt aactatcgtc ttgagtccaa	1740
cccggtaaga cagcacttat cgccactggc agcagccact ggtaacagga ttagcagagc	1800
gaggtatgta ggcgggtgcta cagagttctt gaagtgggtg cctaactacg gctacactag	1860
aaggacagta tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg	1920

29539_0036W01_SL2953936W.txt

tagctcttga	tccggcaaac	aaaccaccgc	tggtagcggg	ggtttttttg	tttgcaagca	1980
gcagattacg	cgcagaaaaa	aaggatctca	agaagatcct	ttgatctttt	ctacgggggc	2040
tgacgctcag	tggaacgaaa	actcacgtta	agggattttg	gtcatgagat	tatcaaaaag	2100
gatcttcacc	tagatccttt	taaattaaaa	atgaagtttt	aaatcaatct	aaagtatata	2160
tgagtaaact	tggtctgaca	gttaccaatg	cttaatcagt	gaggcaccta	tctcagcgat	2220
ctgtctattt	cgttcatcca	tagttgcctg	actccccgtc	gtgtagataa	ctacgatacg	2280
ggagggctta	ccatctggcc	ccagtgtctg	aatgataccg	cgagaccac	gctcaccggc	2340
tccagattta	tcagcaataa	accagccagc	cggaagggcc	gagcgcagaa	gtggctctgc	2400
aactttatcc	gcctccatcc	agtctattaa	ttgttgccgg	gaagctagag	taagtagttc	2460
gccagttaat	agtttgcgca	acgttggtgc	cattgctaca	ggcatcgtgg	tgtcacgctc	2520
gtcgtttggt	atggcttcat	tcagctccgg	ttcccaacga	tcaaggcgag	ttacatgata	2580
ccccatgttg	tgcaaaaaag	cggttagctc	cttcggctct	ccgatcgttg	tcagaagtaa	2640
gttggccgca	gtgttatcac	tcatggttat	ggcagcactg	cataattctc	ttactgtcat	2700
gccatccgta	agatgctttt	ctgtgactgg	tgagtactca	accaagtcac	tctgagaata	2760
gtgtatgcgg	cgaccgagtt	gctcttgccc	ggcgtcaata	cgggataata	ccgcgccaca	2820
tagcagaact	ttaaaagtgc	tcatcattgg	aaaacgttct	tcggggcgaa	aactctcaag	2880
gatcttaccg	ctgttgagat	ccagttcgat	gtaaccact	cgtgcacca	actgatcttc	2940
agcatctttt	actttcacca	gcgtttctgg	gtgagcaaaa	acaggaaggc	aaaatgccgc	3000
aaaaaaggga	ataagggcga	cacggaaatg	ttgaatactc	atactcttcc	tttttcaata	3060
ttattgaagc	atttatcagg	gttattgtct	catgagcgga	tacatatttg	aatgtattta	3120
gaaaaataaa	caaatagggg	ttccgcgcac	atttccccga	aaagtgccac	ctgacgtc	3178

<210> 12
 <211> 111
 <212> DNA
 <213> Homo sapiens

<400> 12	ctacggcgtg	cagtgttca	gccgctacc	cgaccacatg	aagcagcacg	acttcttcaa	60
	gtccgccatg	cccgaaggct	acgtccagga	gcgcaccatc	ttcttcaagg	a	111

<210> 13
 <211> 108
 <212> DNA
 <213> Homo sapiens

<400> 13	ctacggcgtg	cagtgttca	gccgctacc	cgacatgaag	cagcagcact	tcttcaagtc	60
	cgccatgccc	gaaggctacg	tccaggagcg	caccatcttc	ttcaagga		108

<210> 14
 <211> 95
 <212> DNA

<213> Homo sapiens

<400> 14
ctacggcgtg cagtgttca gccgaagcag cagcacttct tcaagtccgc catgcccga 60
ggctacgtcc aggagcgcac catcttcttc aagga 95

<210> 15
<211> 90
<212> DNA
<213> Homo sapiens

<400> 15
ctacggcgtg cagtgttca gccgctacga cttcttcaag tccgccatgc ccgaaggcta 60
cgtccaggag cgcacatct tttcaagga 90

<210> 16
<211> 89
<212> DNA
<213> Homo sapiens

<220>
<221> modified_base
<222> (61)..(61)
<223> a, c, t, g, unknown or other

<400> 16
ctacggcgtg cagtgttca gccgcacgac ttcttccagt ccgccatgcc cgaaggctac 60
ntccaggagc gcacatctt cttcaagga 89

<210> 17
<211> 80
<212> DNA
<213> Homo sapiens

<400> 17
ctacggcgtg cagtgttca gcacttcaag tccgccatgc ccgaaggcta cgtccaggag 60
cgcacatct tttcaagga 80

<210> 18
<211> 78
<212> DNA
<213> Homo sapiens

<400> 18
ctacggcgtg cagcacgact tttcaagtc cgccatgccc gaaggctacg tccaggagcg 60
caccatcttc ttcaagga 78

<210> 19
<211> 78
<212> DNA
<213> Homo sapiens

<400> 19
ctacggcgtg cagcacgact tttcaagtc cgccatgccc gaaggctacg tccaggagcg 60
caccatcttc ttcaagga 78


```

<210> 20
<211> 60
<212> DNA
<213> Homo sapiens

<400> 20
ctacggcgtg cagtgcctgc ccgaaggcta cgtccaggag cgcaccatct tcttcaagga      60

<210> 21
<211> 57
<212> DNA
<213> Homo sapiens

<400> 21
ctacggcgtg cagtgcctca gccgctacgt ccaggagcgc accatcttct tcaagga      57

<210> 22
<211> 51
<212> DNA
<213> Homo sapiens

<400> 22
ctacgccatg cccgaaggct acgtccagga gcgcaccatc ttcttcaagg a      51

<210> 23
<211> 27
<212> DNA
<213> Homo sapiens

<400> 23
ctacggcgtg cagtgccttct tcaagga      27

<210> 24
<211> 113
<212> DNA
<213> Homo sapiens

<400> 24
ctacggcgtg cagtgcctca gccgctaccc cgacaccaca tgaagcagca cgacttcttc      60
aagtcgcgca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa gga      113

<210> 25
<211> 2328
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polynucleotide

<400> 25
tctagagcta gcaccatgga ctacaaagac catgacggtg attataaaga tcatgacatc      60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaagggt gggcattcac      120
cgcggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa      180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc      240
ttcactcatg cgcatattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct      300

```

gtcaaatacc	aagatatgat	tgcggccctg	cccgaagcca	cgcacgaggc	aattgtaggg	360
gtcggtaaac	agtggtcggg	agcgcgagca	cttgaggcgc	tgctgactgt	ggcgggtgag	420
cttagggggc	ctccgctcca	gctcgacacc	gggcagctgc	tgaagatcgc	gaagagaggg	480
ggagtaacag	cggtagaggc	agtgcacgcc	tggcgcaatg	cgctcaccgg	ggcccccttg	540
aacctgaccc	cagaccaggt	agtgcgaatc	gcgaacaata	atgggggaaa	gcaagccctg	600
gaaaccgtgc	aaaggttggt	gccggtcctt	tgtcaagacc	acggccttac	accggagcaa	660
gtcgtggcca	ttgcatccca	cgacggtggc	aaacaggctc	ttgagacggt	tcagagactt	720
ctcccagttc	tctgtcaagc	ccacgggctg	actcccgatc	aagttgtagc	gattgcgtcg	780
aacattggag	ggaacaagc	attggagact	gtccaacggc	tccttcccgt	gttgtgtcaa	840
gcccacggtt	tgacgcctgc	acaagtggtc	gccatcgcca	acaacaacgg	cggtaagcag	900
gcgctggaac	cagtacagcg	cctgctgcct	gtactgtgcc	aggatcatgg	actgacccca	960
gaccaggtag	tcgcaatcgc	gtcaaacgga	gggggaaagc	aagccctgga	aaccgtgcaa	1020
aggttggtgc	cggtcctttg	tcaagaccac	ggccttacac	cggagcaagt	cgtggccatt	1080
gcaaataata	acggtggcaa	acaggctctt	gagacggttc	agagacttct	cccagttctc	1140
tgtcaagccc	acgggctgac	tcccgatcaa	gttgtagcga	ttgcgtcgca	tgacggaggg	1200
aaacaagcat	tggagactgt	ccaacggctc	cttcccgtgt	tgtgtcaagc	ccacggtttg	1260
acgcctgcac	aagtggctgc	catcgctctg	aatggcggcg	gtaagcaggc	gctggaaaca	1320
gtacagcgcc	tgctgcctgt	actgtgccag	gatcatggac	tgaccccaga	ccaggtagtc	1380
gcaatcgcgt	caaacggagg	gggaaagcaa	gccctggaaa	ccgtgcaaag	gttggtgccg	1440
gtcctttgtc	aagaccacgg	ccttacaccg	gagcaagtcg	tggccattgc	atcccacgac	1500
ggtggcaaac	aggctcttga	gacggttcag	agacttctcc	cagttctctg	tcaagcccac	1560
gggctgactc	ccgatcaagt	tgtagcgatt	gcgtcgaaca	ttggagggaa	acaagcattg	1620
gagactgtcc	aacggctcct	tcccgtgttg	tgtcaagccc	acggtttgac	gcctgcacaa	1680
gtggtcgcca	tcgccaacaa	caacggcggt	aagcaggcgc	tggaaacagt	acagcgcctg	1740
ctgcctgtac	tgtgccagga	tcatggactg	accccagacc	aggtagtcgc	aatcgcgtca	1800
catgacgggg	gaaagcaagc	cctggaaacc	gtgcaaaggt	tgttgccggg	cctttgtcaa	1860
gaccacggcc	ttacaccgga	gcaagtcgtg	gccattgcat	cccacgacgg	tggcaaacag	1920
gctcttgaga	cggttcagag	acttctccca	gttctctgtc	aagcccacgg	gctgactccc	1980
gatcaagttg	tagcgattgc	gaataacaat	ggaggggaaac	aagcattgga	gactgtccaa	2040
cggctccttc	ccgtgttggt	tcaagcccac	ggtctgacac	ccgaacaggt	ggtcgccatt	2100
gcttcccacg	acggaggacg	gccagccttg	gagtccatcg	tagcccaatt	gtccaggccc	2160
gatcccgctg	tggctgcgtt	aacgaatgac	catctggtgg	cgttggcatg	tcttggtgga	2220
cgacccgcgc	tcgatgcagt	caaaaagggt	ctgcctcatg	ctcccgcat	gatcaaaaga	2280
accaaccggc	ggattcccga	gagaacttcc	catcgagtcg	cgggatcc		2328

<210> 26

<211> 771

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 26

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

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys

225				230				235				240			
Gln	Ala	His	Gly	Leu 245	Thr	Pro	Asp	Gln	Val 250	Val	Ala	Ile	Ala	Ser 255	Asn
Ile	Gly	Gly	Lys 260	Gln	Ala	Leu	Glu	Thr 265	Val	Gln	Arg	Leu	Leu 270	Pro	Val
Leu	Cys	Gln 275	Ala	His	Gly	Leu	Thr 280	Pro	Ala	Gln	Val	Val 285	Ala	Ile	Ala
Asn	Asn 290	Asn	Gly	Gly	Lys	Gln 295	Ala	Leu	Glu	Thr	Val 300	Gln	Arg	Leu	Leu
Pro 305	Val	Leu	Cys	Gln	Asp 310	His	Gly	Leu	Thr	Pro 315	Asp	Gln	Val	Val	Ala 320
Ile	Ala	Ser	Asn	Gly 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	Glu 350	Gln	Val
Val	Ala	Ile 355	Ala	Asn	Asn	Asn	Gly 360	Gly	Lys	Gln	Ala	Leu 365	Glu	Thr	Val
Gln	Arg 370	Leu	Leu	Pro	Val	Leu 375	Cys	Gln	Ala	His	Gly 380	Leu	Thr	Pro	Asp
Gln 385	Val	Val	Ala	Ile	Ala 390	Ser	His	Asp	Gly	Gly 395	Lys	Gln	Ala	Leu	Glu 400
Thr	Val	Gln	Arg	Leu 405	Leu	Pro	Val	Leu	Cys 410	Gln	Ala	His	Gly	Leu 415	Thr
Pro	Ala	Gln	Val 420	Val	Ala	Ile	Ala	Ser 425	Asn	Gly	Gly	Gly	Lys 430	Gln	Ala
Leu	Glu	Thr 435	Val	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	Gly	Gly	Gly	Lys
Gln 465	Ala	Leu	Glu	Thr	Val 470	Gln	Arg	Leu	Leu	Pro 475	Val	Leu	Cys	Gln	Asp 480
His	Gly	Leu	Thr	Pro 485	Glu	Gln	Val	Val	Ala 490	Ile	Ala	Ser	His	Asp 495	Gly
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

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 530 535 540

Leu Cys Gln Ala His Gly Leu Thr Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 675 680 685

Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Arg Pro Ala
 690 695 700

Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro 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 Ala Leu
 740 745 750

Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His Arg Val
 755 760 765

Ala Gly Ser

770

<210> 27
 <211> 2430
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 27
 tctagagcta gcacatgga ctacaaagac catgacggtg attataaaga tcatgacatc 60
 gattacaagg atgacgatga caagatggcc cccaagaaga agaggaaggt gggcattcac 120
 cgcgggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa 180
 atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc 240
 ttcactcatg cgcataattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct 300
 gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg 360
 gtcggtaaag agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag 420
 cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg 480
 ggagtaacag cggtagaggc agtgcacgcc tggcgcaatg cgctcaccgg ggcccccttg 540
 aacctgacct cagaccaggt agtcgcaatc gcgaacaata atgggggaaa gcaagccctg 600
 gaaaccgtgc aaaggttgtt gccggtcctt tgtcaagacc acggccttac accggagcaa 660
 gtcgtggcca ttgcatccca cgacggtggc aaacaggctc ttgagacggt tcagagactt 720
 ctcccagttc tctgtcaagc ccacgggctg actcccgatc aagttgtagc gattgcgtcg 780
 aacattggag ggaacaagc attggagact gtccaacggc tccttcccgt gttgtgtcaa 840
 gccacgggtt tgacgcctgc acaagtggtc gccatcgcca acaacaacgg cggtaagcag 900
 gcgctggaaa cagtacagcg cctgctgcct gtactgtgcc aggatcatgg actgacccca 960
 gaccaggtag tcgcaatcgc gtcaaacgga gggggaaagc aagccctgga aaccgtgcaa 1020
 aggttggtgc cggtcctttg tcaagaccac ggccttacac cggagcaagt cgtggccatt 1080
 gcaaataata acggtggcaa acaggctctt gagacggttc agagacttct cccagttctc 1140
 tgtcaagccc acgggctgac tcccgatcaa gttgtagcga ttgcgtcgca tgacggaggg 1200
 aaacaagcat tggagactgt ccaacggctc cttcccgtgt tgtgtcaagc ccacggtttg 1260
 acgcctgcac aagtggctgc catcgctcgc aatggcggcg gtaagcaggc gctggaaaca 1320
 gtacagcgcc tgctgcctgt actgtgccag gatcatggac tgaccccaga ccaggtagtc 1380
 gcaatcgcgt caaacggagg gggaaagcaa gccctggaaa ccgtgcaaag gttgttgccg 1440
 gtcctttgtc aagaccacgg ccttacaccg gagcaagtcg tggccattgc atcccacgac 1500
 ggtggcaaag aggtctctga gacggttcag agacttctcc cagttctctg tcaagccac 1560
 gggctgactc ccgatcaagt ttagcgaatt gcgtcgaaca ttggagggaa acaagcattg 1620
 gagactgtcc aacggctcct tcccgtgttg tgtcaagccc acggtttgac gcctgcacaa 1680

```

gtggtcgcca tcgccaacaa caacggcggg aagcaggcgc tggaaacagt acagcgccgtg      1740
ctgcctgtac tgtgccagga tcatggactg accccagacc aggtagtcgc aatcgcgctca      1800
catgacgggg gaaagcaagc cctggaaacc gtgcaaaggt tgttgccggg cctttgtcaa      1860
gaccacggcc ttacaccgga gcaagtcgtg gccattgcat cccacgacgg tggcaaacag      1920
gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc      1980
gatcaagttg tagcgattgc gaataacaat ggagggaaac aagcattgga gactgtccaa      2040
cggctccttc ccgtgttgtg tcaagcccac ggtttgacgc ctgcacaagt ggtcgccatc      2100
gccagccatg atggcggtaa gcaggcgctg gaaacagtac agcgccctgct gcctgtactg      2160
tgccaggatc atggactgac acccgaacag gtggtcgcca ttgcttctaa tgggggagga      2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatcccgc gttggctgcg      2280
ttaacgaatg accatctggt ggcgttggca tgtcttggtg gacgaccgc gctcgatgca      2340
gtcaaaaagg gtctgcctca tgctcccgca ttgatcaaaa gaaccaaccg gcggattccc      2400
gagagaactt cccatcgagt cgcgggatcc      2430

```

<210> 28

<211> 805

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 28

```

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1           5           10          15

```

```

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
          20          25          30

```

```

Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr Leu Gly Tyr
          35          40          45

```

```

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
          50          55          60

```

```

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
65          70          75          80

```

```

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
          85          90          95

```

```

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
          100         105         110

```

```

Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
          115         120         125

```

Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp
 130 135 140
 Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
 145 150 155 160
 Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
 165 170 175
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys
 180 185 190
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 195 200 205
 His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly
 210 215 220
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 225 230 235 240
 Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 245 250 255
 Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 260 265 270
 Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285
 Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 305 310 315 320
 Ile Ala Ser Asn Gly 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 Glu 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 Ala 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 Leu₄₀₅ Leu Pro Val Leu Cys₄₁₀ Gln Ala His Gly Leu₄₁₅ Thr
 Pro Ala Gln Val₄₂₀ Val Ala Ile Ala Ser₄₂₅ Asn Gly Gly Gly Lys₄₃₀ Gln Ala
 Leu Glu Thr Val₄₃₅ Gln Arg Leu Leu₄₄₀ Pro Val Leu Cys Gln₄₄₅ Asp His Gly
 Leu Thr₄₅₀ Pro Asp Gln Val Val₄₅₅ Ala Ile Ala Ser Asn₄₆₀ Gly Gly Gly Lys
 Gln Ala Leu Glu Thr Val₄₇₀ Gln Arg Leu Leu₄₇₅ Pro Val Leu Cys Gln Asp₄₈₀
 His Gly Leu Thr Pro₄₈₅ Glu Gln Val Val Ala₄₉₀ Ile Ala Ser His Asp₄₉₅ Gly
 Gly Lys Gln Ala₅₀₀ Leu Glu Thr Val Gln₅₀₅ Arg Leu Leu Pro Val₅₁₀ Leu Cys
 Gln Ala His₅₁₅ Gly Leu Thr Pro Asp₅₂₀ Gln Val Val Ala Ile₅₂₅ Ala Ser Asn
 Ile Gly₅₃₀ Gly Lys Gln Ala Leu₅₃₅ Glu Thr Val Gln Arg₅₄₀ Leu Leu Pro Val
 Leu Cys Gln Ala His Gly₅₅₀ Leu Thr Pro Ala Gln₅₅₅ Val Val Ala Ile Ala₅₆₀
 Asn Asn Asn Gly Gly₅₆₅ Lys Gln Ala Leu Glu₅₇₀ Thr Val Gln Arg Leu Leu₅₇₅
 Pro Val Leu Cys₅₈₀ Gln Asp His Gly Leu₅₈₅ Thr Pro Asp Gln Val₅₉₀ Val Ala
 Ile Ala Ser₅₉₅ His Asp Gly Gly Lys₆₀₀ Gln Ala Leu Glu Thr Val₆₀₅ Gln Arg
 Leu Leu Pro Val Leu Cys Gln₆₁₅ Asp His Gly Leu Thr Pro Glu Gln Val₆₂₀
 Val Ala Ile Ala Ser His₆₃₀ Asp Gly Gly Lys Gln₆₃₅ Ala Leu Glu Thr Val₆₄₀
 Gln Arg Leu Leu Pro₆₄₅ Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp₆₅₅
 Gln Val Val Ala₆₆₀ Ile Ala Asn Asn Asn₆₆₅ Gly Gly Lys Gln Ala Leu Glu₆₇₀

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr
675 680 685

Pro Ala 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 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Arg
725 730 735

Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
740 745 750

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
755 760 765

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
770 775 780

Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
785 790 795 800

Arg Val Ala Gly Ser
805

<210> 29

<211> 2430

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 29

tctagagcta gcacatgga ctacaaagac catgacggtg attataaaga tcatgacatc	60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaaggt gggcattcac	120
cgcggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa	180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc	240
ttcactcatg cgcatttgt cgcgctttca cagcaccctg cggcgcttg gacggtggct	300
gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg	360
gtcggtaaac agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag	420
cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg	480
ggagtaacag cggtagaggc agtcgacgcc tggcgcaatg cgctcaccgg ggcccccttg	540
aacctgaccc cagaccaggt agtcgcaatc gcgtcaaacg gagggggaaa gcaagccctg	600

29539_0036W01_SL2953936W.txt

gaaaccgtgc aaaggttggt gccggtcctt tgtcaagacc acggccttac accggagcaa	660
gtcgtggcca ttgcaaataa taacggtggc aaacaggctc ttgagacggt tcagagactt	720
ctcccagttc tctgtcaagc ccacgggctg actcccgatc aagttgtagc gattgcgtcg	780
aacattggag ggaaacaagc attggagact gtccaacggc tccttcccgt gttgtgtcaa	840
gcccacggtt tgacgcctgc acaagtggtc gccatcgctt ccaatattgg cggtaagcag	900
gcgctggaaa cagtacagcg cctgctgcct gtactgtgcc aggatcatgg actgacccca	960
gaccaggtag tcgcaatcgc gaacaataat gggggaaagc aagccctgga aaccgtgcaa	1020
aggttggtgc cggtcctttg tcaagaccac ggccttacac cggagcaagt cgtggccatt	1080
gcaagcaaca tcggtggcaa acaggctctt gagacggttc agagacttct cccagttctc	1140
tgtcaagccc acgggctgac tcccgatcaa gttgtagcga ttgcgtcgaa cattggaggg	1200
aaacaagcat tggagactgt ccaacggctc cttcccgtgt tgtgtcaagc ccacggtttg	1260
acgcctgcac aagtggctgc catcgccaac aacaacggcg gtaagcaggc gctggaaaca	1320
gtacagcgcc tgctgcctgt actgtgccag gatcatggac tgaccccaga ccaggtagtc	1380
gcaatcgcgt caaacggagg gggaaagcaa gccctggaaa cgtgcaaag gttgttgccg	1440
gtcctttgtc aagaccacgg ccttacaccg gagcaagtcg tggccattgc atcccacgac	1500
ggtggcaaac aggctcttga gacggttcag agacttctcc cagttctctg tcaagcccac	1560
gggctgactc ccgatcaagt tgtagcgatt gcgaataaca atggagggaa acaagcattg	1620
gagactgtcc aacggctcct tcccgtgttg tgtcaagccc acggtttgac gcctgcacaa	1680
gtggtcgcca tcgcctcgaa tggcggcggt aagcaggcgc tggaaacagt acagcgctg	1740
ctgcctgtac tgtgccagga tcatggactg accccagacc aggtagtcgc aatcgcgaac	1800
aataatgggg gaaagcaagc cctggaaacc gtgcaaaggt tgttgccggt cctttgtcaa	1860
gaccacggcc ttacaccgga gcaagtcgtg gccattgcat cccacgacgg tggcaaacag	1920
gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc	1980
gatcaagttg tagcgattgc gtccaacggg ggagggaaac aagcattgga gactgtccaa	2040
cggctccttc ccgtgttgtg tcaagcccac ggtttgacgc ctgcacaagt ggtcgccatc	2100
gccaacaaca acggcggtaa gcaggcgctg gaaacagtag agcgctgct gcctgtactg	2160
tgccaggatc atggactgac acccgaacag gtggctcgcca ttgcttccca cgacggagga	2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatcccgc gttggctgcg	2280
ttaacgaatg accatctggt ggcgttgga tgtcttggtg gacgaccgc gctcgatgca	2340
gtcaaaaagg gtctgcctca tgctcccgca ttgatcaaaa gaaccaaccg gcggattccc	2400
gagagaactt cccatcgagt cgcgggatcc	2430

<210> 30

<211> 805

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 30

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

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 260 265 270

Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285

Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu Gln Val
 340 345 350

Val Ala Ile Ala Ser Asn Ile 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 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 Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr Pro Ala 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 Asn Asn Asn 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 675 680 685
 Pro Ala 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 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Arg
 725 730 735
 Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
 740 745 750
 Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
 755 760 765
 Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
 770 775 780
 Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
 785 790 795 800

Arg Val Ala Gly Ser
805

<210> 31
<211> 2430
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

```

<400> 31
tctagagcta gcaccatgga ctacaaagac catgacggtg attataaaga tcatgacatc      60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaagggt gggcattcac      120
cgcggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa      180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc      240
ttcactcatg cgcatattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct      300
gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg      360
gtcggtaaac agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag      420
cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg      480
ggagtaacag cggtagaggc agtgcacgcc tggcgcaatg cgctcaccgg ggcccccttg      540
aacctgaccc cagaccaggt agtcgcaatc gcgaacaata atgggggaaa gcaagccctg      600
gaaaccgtgc aaaggttggt gccggtcctt tgtcaagacc acggccttac accggagcaa      660
gtcgtggcca ttgcaagcaa catcgggtggc aaacaggctc ttgagacggt tcagagactt      720
ctcccagttc tctgtcaagc ccacgggctg actcccgatc aagttgtagc gattgcgtcg      780
aacattggag ggaaacaagc attggagact gtccaacggc tccttcccgt gttgtgtcaa      840
gcccacgggt tgacgcctgc acaagtggtc gccatcgcca acaacaacgg cggtaaagcag      900
gcgctggaac cagtacagcg cctgctgcct gtactgtgcc aggatcatgg actgacccca      960
gaccaggtag tcgcaatcgc gtcgaacatt gggggaaagc aagccctgga aaccgtgcaa     1020
aggttggtgc cggtcctttg tcaagaccac ggccttacac cggagcaagt cgtggccatt     1080
gcaagcaaca tcggtggcaa acaggctctt gagacggttc agagacttct cccagttctc     1140
tgtcaagccc acgggctgac tcccgatcaa gttgtagcga ttgcgaataa caatggaggg     1200
aaacaagcat tggagactgt ccaacggctc cttcccgtgt tgtgtcaagc ccacggtttg     1260
acgcctgcac aagtggtcgc catcgcctcg aatggcggcg gtaagcaggc gctggaacaa     1320
gtacagcgcc tgctgcctgt actgtgccag gatcatggac tgaccccaaga ccaggtagtc     1380
gcaatcgcgt cacatgacgg gggaaagcaa gccctggaaa ccgtgcaaag gttgttgccg     1440
gtcctttgtc aagaccacgg ccttacaccg gagcaagtcg tggccattgc aaataataac     1500
ggtaggcaaac aggtccttga gacggttcag agacttctcc cagttctctg tcaagcccac     1560
gggctgactc ccgatcaagt tgtagcgatt gcgtccaacg gtggagggaa acaagcattg     1620

```

```

gagactgtcc aacggctcct tcccgtgttg tgtcaagccc acggtttgac gcctgcacaa 1680
gtggtcgcca tcgccaacaa caacggcggg aagcaggcgc tggaaacagt acagcgcttg 1740
ctgcctgtac tgtgccagga tcatggactg accccagacc aggtagtcgc aatcgcgta 1800
catgacgggg gaaagcaagc cctggaaacc gtgcaaaggt tgttgccggg cctttgtcaa 1860
gaccacggcc ttacaccgga gcaagtcgtg gccattgcaa gcaatggggg tggcaaacag 1920
gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc 1980
gatcaagttg tagcgattgc gaataacaat ggagggaac aagcattgga gactgtccaa 2040
cggctccttc ccgtgttgtg tcaagcccac ggtttgacgc ctgcacaagt ggtcgccatc 2100
gccagccatg atggcggtaa gcaggcgtg gaaacagtac agcgctgct gcctgtactg 2160
tgccaggatc atggactgac acccgaacag gtggtcgcca ttgcttctaa tgggggagga 2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatccgc gttggctgcg 2280
ttaacgaatg accatctggg ggcgttgga tgtcttggtg gacgaccgc gtcgatgca 2340
gtcaaaaagg gtctgcctca tgctccgca ttgatcaaaa gaaccaaccg gcggattccc 2400
gagagaactt cccatcgagt cgcgggatcc 2430

```

<210> 32
 <211> 805
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polypeptide

```

<400> 32
Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1          5          10         15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
          20          25          30

Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr Leu Gly Tyr
          35          40          45

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
          50          55          60

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
65          70          75          80

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
          85          90          95

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
          100         105         110

```


Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
 115 120 125
 Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu Asp
 130 135 140
 Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
 145 150 155 160
 Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
 165 170 175
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys
 180 185 190
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 195 200 205
 His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly
 210 215 220
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 225 230 235 240
 Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 245 250 255
 Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 260 265 270
 Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285
 Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu Gln Val
 340 345 350
 Val Ala Ile Ala Ser Asn Ile 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 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 Ala His Gly Leu Thr
 405 410 415
 Pro Ala 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 Glu 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 Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 515 520 525
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 530 535 540
 Leu Cys Gln Ala His Gly Leu Thr Pro Ala 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 Glu Gln Val
 610 615 620
 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 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 Ala His Gly Leu Thr
675 680 685

Pro Ala 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 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Arg
725 730 735

Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
740 745 750

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
755 760 765

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
770 775 780

Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
785 790 795 800

Arg Val Ala Gly Ser
805

<210> 33

<211> 2328

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 33

tctagagcta gcacatgga ctacaaagac catgacggtg attataaaga tcatgacatc	60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaaggt gggcattcac	120
cgcggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa	180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc	240
ttcactcatg cgcatattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct	300
gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg	360
gtcggtaaac agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag	420
cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg	480
ggagtaacag cggtagaggc agtgcacgcc tggcgcaatg cgctcaccgg ggcccccttg	540

aacctgaccc	cagaccaggt	agtcgcaatc	gcgtcacatg	acgggggaaa	gcaagccctg	600
gaaaccgtgc	aaaggttggt	gccggtcctt	tgtcaagacc	acggccttac	accggagcaa	660
gtcgtggcca	ttgcaaataa	taacggtggc	aaacaggctc	ttgagacggt	tcagagactt	720
ctcccagttc	tctgtcaagc	ccacgggctg	actcccgatc	aagttgtagc	gattgcgtcg	780
aacattggag	ggaaacaagc	attggagact	gtccaacggc	tccttcccgt	gttgtgtcaa	840
gcccacgggt	tgacgcctgc	acaagtggtc	gccatcgcca	acaacaacgg	cggtaagcag	900
gcgctggaaa	cagtacagcg	cctgctgcct	gtactgtgcc	aggatcatgg	actgacccca	960
gaccaggtag	tcgcaatcgc	gtcacatgac	gggggaaagc	aagccctgga	aaccgtgcaa	1020
aggttggtgc	cggtcctttg	tcaagaccac	ggccttacac	cggagcaagt	cgtggccatt	1080
gcaagcaatg	ggggtggcaa	acaggctctt	gagacggttc	agagacttct	cccagttctc	1140
tgtcaagccc	acgggctgac	tcccgatcaa	gttgtagcga	ttgcgaataa	caatggaggg	1200
aaacaagcat	tggagactgt	ccaacggctc	cttcccgtgt	tgtgtcaagc	ccacggtttg	1260
acgcctgcac	aagtggtcgc	catcgctctc	aatattggcg	gtaagcaggc	gctggaaaca	1320
gtacagcgcc	tgctgcctgt	actgtgccag	gatcatggac	tgaccccaga	ccaggtagtc	1380
gcaatcgctg	cgaacattgg	gggaaagcaa	gccctggaaa	ccgtgcaaag	gttggtgccg	1440
gtcctttgtc	aagaccacgg	ccttacaccg	gagcaagtcg	tggccattgc	aaataataac	1500
ggtggcaaac	aggctcttga	gacggttcag	agacttctcc	cagttctctg	tcaagcccac	1560
gggctgactc	ccgatcaagt	tgtagcgatt	gcgaataaca	atggagggaa	acaagcattg	1620
gagactgtcc	aacggctcct	tcccgtgttg	tgtcaagccc	acggtttgac	gcctgcacaa	1680
gtggctgcca	tcgccaacaa	caacggcggt	aagcaggcgc	tggaaacagt	acagcgctcg	1740
ctgcctgtac	tgtgccagga	tcatggactg	accccagacc	aggtagtcgc	aatcgcgta	1800
catgacgggg	gaaagcaagc	cctggaaacc	gtgcaaaggt	tgttgccggt	cctttgtcaa	1860
gaccacggcc	ttacaccgga	gcaagtcgtg	gccattgcaa	gcaacatcgg	tggcaaacag	1920
gctcttgaga	cggttcagag	acttctccca	gttctctgtc	aagcccacgg	gctgactccc	1980
gatcaagttg	tagcgattgc	gtccaacggc	ggagggaaac	aagcattgga	gactgtccaa	2040
cggctccttc	ccgtgttggt	tcaagcccac	ggtctgacac	ccgaacaggc	ggtcgccatt	2100
gcttcccacg	acggaggacg	gccagccttg	gagtccatcg	tagcccaatt	gtccaggccc	2160
gatcccgctg	tggctgcgtt	aacgaatgac	catctggtgg	cgttggcatg	tcttggtgga	2220
cgacccgcgc	tcgatgcagt	caaaaagggt	ctgcctcatg	ctcccgcatg	gatcaaaaga	2280
accaaccggc	ggattcccgga	gagaacttcc	catcgagtcg	cgggatcc		2328

<210> 34

<211> 771

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polypeptide

<400> 34

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
 1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
 20 25 30

Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr Leu Gly Tyr
 35 40 45

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
 50 55 60

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
 65 70 75 80

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
 85 90 95

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
 100 105 110

Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
 115 120 125

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

Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
 145 150 155 160

Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
 165 170 175

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 180 185 190

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 195 200 205

His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly
 210 215 220

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 225 230 235 240

Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 245 250 255

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

260

265

270

Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285

Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu 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 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 Ala His Gly Leu Thr
 405 410 415

Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 675 680 685

Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Arg Pro Ala
 690 695 700

Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro 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 Ala Leu
 740 745 750

Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His Arg Val
 755 760 765

Ala Gly Ser
 770

<210> 35

<211> 2430

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 35

tctagagcta gcacatgga ctacaaagac catgacggtg attataaaga tcatgacatc	60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaaggt gggcattcac	120
cgcggggtac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa	180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc	240
ttcactcatg cgcatattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct	300
gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg	360
gtcggtaaac agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag	420
cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg	480
ggagtaacag cggtagaggc agtgcacgcc tggcgcaatg cgctcaccgg ggcccccttg	540
aacctgaccc cagaccaggt agtcgcaatc gcgtcacatg acgggggaaa gcaagccctg	600
gaaaccgtgc aaaggttggt gccggtcctt tgtcaagacc acggccttac accggagcaa	660
gtcgtggcca ttgcaaataa taacggtggc aaacaggctc ttgagacggt tcagagactt	720
ctcccagttc tctgtcaagc ccacgggctg actcccgatc aagttgtagc gattgcgtcg	780
aacattggag ggaacaagc attggagact gtccaacggc tccttcccgt gttgtgtcaa	840
gccacgggtt tgacgcctgc acaagtggtc gccatcgcca acaacaacgg cggtaagcag	900
gcgctggaaa cagtacagcg cctgctgcct gtactgtgcc aggatcatgg actgacccca	960
gaccaggtag tcgcaatcgc gtcacatgac gggggaaagc aagccctgga aaccgtgcaa	1020
aggttgttgc cggtcctttg tcaagaccac ggccttacac cggagcaagt cgtggccatt	1080
gcaagcaatg ggggtggcaa acaggctctt gagacggttc agagacttct cccagttctc	1140
tgtcaagccc acgggctgac tcccgatcaa gttgtagcga ttgcgaataa caatggaggg	1200
aaacaagcat tggagactgt ccaacggctc cttcccgtgt tgtgtcaagc ccacggtttg	1260
acgcctgcac aagtggtcgc catcgctcct aatattggcg gtaagcaggc gctggaaaca	1320
gtacagcgcc tgctgcctgt actgtgccag gatcatggac tgaccccaga ccaggtagtc	1380
gcaatcgcgt cgaacattgg gggaaagcaa gccctggaaa ccgtgcaaag gttgttgccg	1440
gtcctttgtc aagaccacgg ccttacaccg gagcaagtcg tggccattgc aaataataac	1500
ggtggcaaac aggctcttga gacggttcag agacttctcc cagttctctg tcaagcccac	1560
gggctgactc ccgatcaagt tgtagcgatt gcgaataaca atggagggaa acaagcattg	1620
gagactgtcc aacggctcct tcccgtgttg tgtcaagccc acggtttgac gcctgcacaa	1680
gtggtcgcca tcgccaacaa caacggcggt aagcaggcgc tggaaacagt acagcgcttg	1740
ctgcctgtac tgtgccagga tcatggactg accccagacc aggtagtcgc aatcgcgta	1800
catgacgggg gaaagcaagc cctggaaacc gtgcaaaggt tgttgccggg cctttgtcaa	1860
gaccacggcc ttacaccgga gcaagtcgtg gccattgcaa gcaacatcgg tggcaaacag	1920


```

gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc 1980
gatcaagttg tagcgattgc gtccaacggg ggagggaaac aagcattgga gactgtccaa 2040
cggctccttc ccgtgttggt tcaagcccac ggtttgacgc ctgcacaagt ggtcgccatc 2100
gccagccatg atggcggtaa gcaggcgctg gaaacagtac agcgcttgct gcctgtactg 2160
tgccaggatc atggactgac acccgaacag gtggtcgcca ttgctaataa taacggagga 2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatcccgc gttggctgcg 2280
ttaacgaatg accatctggt ggcgttgga tgtcttggtg gacgaccgc gctcgatgca 2340
gtcaaaaagg gtctgcctca tgctcccgca ttgatcaaaa gaaccaaccg gcggattccc 2400
gagagaactt cccatcgagt cgcgggatcc 2430

```

<210> 36

<211> 805

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 36

```

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1          5          10          15

```

```

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
          20          25          30

```

```

Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr Leu Gly Tyr
          35          40          45

```

```

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
50          55          60

```

```

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
65          70          75          80

```

```

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
          85          90          95

```

```

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
          100          105          110

```

```

Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
          115          120          125

```

```

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

```

```

Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
          145          150          155          160

```

Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
 165 170 175
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 180 185 190
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 195 200 205
 His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly
 210 215 220
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 225 230 235 240
 Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 245 250 255
 Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 260 265 270
 Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285
 Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu 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 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 Ala His Gly Leu Thr
 405 410 415
 Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 675 680 685
 Pro Ala 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 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Arg
725 730 735

Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
740 745 750

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
755 760 765

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
770 775 780

Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
785 790 795 800

Arg Val Ala Gly Ser
805

<210> 37
<211> 2430
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 37
tctagagcta gcacatgga ctacaaagac catgacggtg attataaaga tcatgacatc 60
gattacaagg atgacgatga caagatggcc cccaagaaga agaggaaggt gggcattcac 120
cgcggggttac ctatggtgga cttgaggaca ctcggttatt cgcaacagca acaggagaaa 180
atcaagccta aggtcaggag caccgtcgcg caacaccacg aggcgcttgt ggggcatggc 240
ttcactcatg cgcatattgt cgcgctttca cagcaccctg cggcgcttgg gacggtggct 300
gtcaaatacc aagatatgat tgcggccctg cccgaagcca cgcacgaggc aattgtaggg 360
gtcggtaaac agtggtcggg agcgcgagca cttgaggcgc tgctgactgt ggcgggtgag 420
cttagggggc ctccgctcca gctcgacacc gggcagctgc tgaagatcgc gaagagaggg 480
ggagtaacag cggtagaggc agtcacgcc tggcgcaatg cgctcaccgg ggcccccttg 540
aacctgaccc cagaccaggt agtcgcaatc gcgtcaaacg gagggggaaa gcaagccctg 600
gaaaccgtgc aaaggttggt gccggtcctt tgtcaagacc acggccttac accggagcaa 660
gtcgtggcca ttgcaaataa taacggtggc aaacaggctc ttgagacggt tcagagactt 720
ctcccagttc tctgtcaagc ccacgggctg actcccgatc aagttgtagc gattgcgtcc 780
aacggtggag ggaacaagc attggagact gtccaacggc tccttcccgt gttgtgtcaa 840

29539_0036W01_SL2953936W.txt

```

gccacggtt tgacgcctgc acaagtgggc gccatcgcca acaacaacgg cggtaagcag    900
gcgctggaaa cagtacagcg cctgctgcct gtactgtgcc aggatcatgg actgacccca    960
gaccaggtag tcgcaatcgc gtcacatgac gggggaaagc aagccctgga aaccgtgcaa   1020
aggttgttgc cggtcctttg tcaagaccac ggccttacac cggagcaagt cgtggccatt   1080
gcatcccacg acggtggcaa acaggctctt gagacggttc agagacttct cccagttctc   1140
tgtcaagccc acgggctgac tcccgatcaa gttgtagcga ttgcgtcgca tgacggaggg   1200
aaacaagcat tggagactgt ccaacggctc cttcccggtg tgtgtcaagc ccacggtttg   1260
acgcctgcac aagtggctgc catcgccagc catgatggcg gtaagcaggc gctggaaaca   1320
gtacagcgcc tgctgcctgt actgtgccag gatcatggac tgaccccaga ccaggtagtc   1380
gcaatcgcgt cgaacattgg gggaaagcaa gccctggaaa ccgtgcaaag gttgttgccg   1440
gtcctttgtc aagaccacgg ccttacaccg gagcaagtcg tggccattgc aaataataac   1500
ggtggcaaac aggctcttga gacggttcag agacttctcc cagttctctg tcaagccac   1560
gggctgactc ccgatcaagt tgtagcgatt gcgaataaca atggagggaa acaagcattg   1620
gagactgtcc aacggctcct tcccggtgtg tgtcaagccc acggtttgac gcctgcacaa   1680
gtggtcgcca tcgcctccaa tattggcggg aagcaggcgc tggaaacagt acagcgcctg   1740
ctgcctgtac tgtgccagga tcatggactg accccagacc aggtagtcgc aatcgcgtca   1800
aacggagggg gaaagcaagc cctggaaacc gtgcaaagggt tgttgccggg cttttgtcaa   1860
gaccacggcc ttacaccgga gcaagtcgtg gccattgcaa ataataacgg tggcaaacag   1920
gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc   1980
gatcaagttg tagcgattgc gtccaacggg ggagggaaac aagcattgga gactgtccaa   2040
cggctccttc ccgtgttgtg tcaagccac ggtttgacgc ctgcacaagt ggctgccatc   2100
gcctcgaatg gcggcggtaa gcaggcgctg gaaacagtac agcgcctgct gcctgtactg   2160
tgccaggatc atggactgac acccgaacag gtggtcgcca ttgctaataa taacggagga   2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatcccgc gttggctgcg   2280
ttaacgaatg accatctggt ggcgttgga tgtcttggtg gacgaccgc gctcgatgca   2340
gtcaaaaagg gtctgcctca tgctcccgca ttgatcaaaa gaaccaaccg gcggattccc   2400
gagagaactt cccatcgagt cgcgggatcc                                2430

```

<210> 38

<211> 805

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 38

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
 1 5 10 15

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

Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 405 410 415
 Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr Pro Ala 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 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 Glu 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 Ala 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 Ala His Gly Leu Thr
675 680 685

Pro Ala 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 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Arg
725 730 735

Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
740 745 750

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
755 760 765

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
770 775 780

Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
785 790 795 800

Arg Val Ala Gly Ser
805

<210> 39
<211> 2430
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 39

tctagagcta	gcacatgga	ctacaaagac	catgacggtg	attataaaga	tcatgacatc	60
gattacaagg	atgacgatga	caagatggcc	cccaagaaga	agaggaaggt	gggcattcac	120
cgcggggtac	ctatggtgga	cttgaggaca	ctcggttatt	cgcaacagca	acaggagaaa	180
atcaagccta	aggtcaggag	caccgtcgcg	caacaccacg	aggcgcttgt	ggggcatggc	240
ttcactcatg	cgcataattgt	cgcgctttca	cagcaccctg	cggcgcttgg	gacggtggct	300
gtcaaatacc	aagatatgat	tgcggccctg	cccgaagcca	cgcacgaggc	aattgtaggg	360
gtcggtaaac	agtggtcggg	agcgcgagca	cttgaggcgc	tgctgactgt	ggcgggtgag	420
cttagggggc	ctccgctcca	gctcgacacc	gggcagctgc	tgaagatcgc	gaagagaggg	480
ggagtaacag	cggtagaggc	agtgcacgcc	tggcgcaatg	cgctcaccgg	ggcccccttg	540
aacctgacct	cagaccaggt	agtcgcaatc	gcgaacaata	atgggggaaa	gcaagccctg	600
gaaaccgtgc	aaaggttggt	gccggtcctt	tgtcaagacc	acggccttac	accggagcaa	660
gtcgtggcca	ttgcaagcaa	tgggggtggc	aaacaggctc	ttgagacggt	tcagagactt	720
ctcccagttc	tctgtcaagc	ccacgggctg	actcccgatc	aagttgtagc	gattgcgaat	780
aacaatggag	ggaaacaagc	attggagact	gtccaacggc	tccttcccgt	gttgtgtcaa	840
gccacagggt	tgacgcctgc	acaagtggtc	gccatcgcca	gccatgatgg	cggtaagcag	900
gcgctggaaa	cagtacagcg	cctgctgcct	gtactgtgcc	aggatcatgg	actgacccca	960
gaccaggtag	tcgcaatcgc	gtcacatgac	gggggaaagc	aagccctgga	aaccgtgcaa	1020
aggttggtgc	cggtcctttg	tcaagaccac	ggccttacac	cggagcaagt	cgtggccatt	1080
gcatcccacg	acggtggcaa	acaggctctt	gagacggttc	agagacttct	cccagttctc	1140
tgtcaagccc	acgggctgac	tcccgatcaa	gttgtagcga	ttgcgtcgca	tgacggaggg	1200
aaacaagcat	tggagactgt	ccaacggctc	cttcccgtgt	tgtgtcaagc	ccacggtttg	1260
acgcctgcac	aagtggtcgc	catgcctcc	aatattggcg	gtaagcaggc	gctggaaaca	1320
gtacagcgcc	tgctgcctgt	actgtgccag	gatcatggac	tgaccccgaa	ccaggtagtc	1380
gcaatcgcg	acaataatgg	gggaaagcaa	gccctggaaa	ccgtgcaaag	gttggtgccg	1440
gtcctttgtc	aagaccacgg	ccttacaccg	gagcaagtcg	tggccattgc	aaataataac	1500
ggtggcaaac	aggctcttga	gacggttcag	agacttctcc	cagttctctg	tcaagccac	1560
gggctgactc	ccgatcaagt	tgtagcgatt	gcgtcgaaca	ttggagggaa	acaagcattg	1620
gagactgtcc	aacggctcct	tcccgtgttg	tgtcaagccc	acggtttgac	gcctgcacaa	1680
gtggtcgcca	tcgcctcgaa	tggcggcggt	aagcaggcgc	tggaaacagt	acagcgccctg	1740
ctgcctgtac	tgtgccagga	tcatggactg	accccagacc	aggtagtcgc	aatcgcgaac	1800
aataatgggg	gaaagcaagc	cctggaaacc	gtgcaaagggt	tgttgccgggt	cctttgtcaa	1860

gaccacggcc ttacaccgga gcaagtcgtg gccattgcaa gcaatggggg tggcaaacag 1920
gctcttgaga cggttcagag acttctccca gttctctgtc aagcccacgg gctgactccc 1980
gatcaagttg tagcgattgc gtccaacggt ggaggggaaac aagcattgga gactgtccaa 2040
cggctccttc ccgtgttgtg tcaagcccac ggtttgacgc ctgcacaagt ggtcgccatc 2100
gccaacaaca acggcggtta gcaggcgctg gaaacagtac agcgctgtgt gcctgtactg 2160
tgccaggatc atggactgac acccgaacag gtggtcgcca ttgcttccca cgacggagga 2220
cggccagcct tggagtccat cgtagcccaa ttgtccaggc ccgatcccgc gttggctgcg 2280
ttaacgaatg accatctggt ggcgttgga tgtcttggtg gacgaccgc gctcgatgca 2340
gtcaaaaagg gtctgcctca tgctcccgca ttgatcaaaa gaaccaaccg gcggattccc 2400
gagagaactt cccatcgagt cgcgggatcc 2430

<210> 40

<211> 805

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 40

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp
1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys Arg Lys Val
20 25 30

Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr Leu Gly Tyr
35 40 45

Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val
50 55 60

Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His Ala His
65 70 75 80

Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val Ala Val
85 90 95

Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His Glu Ala
100 105 110

Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu Glu Ala
115 120 125

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

Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Val
 145 150 155 160
 Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
 165 170 175
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys
 180 185 190
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 195 200 205
 His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly
 210 215 220
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 225 230 235 240
 Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn
 245 250 255
 Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 260 265 270
 Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 275 280 285
 Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 290 295 300
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp 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 Glu 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 Ala 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 Ala His Gly Leu Thr
 405 410 415

Pro Ala 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 Glu 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 Ala 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 Ala His Gly Leu Thr Pro Ala 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 Asn Asn Asn 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 Glu Gln Val
 610 615 620

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 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 Ala His Gly Leu Thr
 675 680 685

Pro Ala 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 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Arg
 725 730 735

Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
 740 745 750

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
 755 760 765

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro
 770 775 780

Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His
 785 790 795 800

Arg Val Ala Gly Ser
 805

<210> 41
 <211> 16
 <212> DNA
 <213> Homo sapiens

<400> 41
 gcagtgccttc agccgc

16

<210> 42
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 primer

<400> 42
 tctagagaag acaagaacct gacc

24

<210> 43
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 primer

<220>
 <223> 5'-biotin

<400> 43

29539_0036W01_SL2953936W.txt
ggatccggtc tcttaaggcc gtgg 24

<210> 44
<211> 16
<212> DNA
<213> Homo sapiens

<400> 44
gcagtgcttc agccgc 16

<210> 45
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 45
tctaacatc 9

<210> 46
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 46
tcccacgac 9

<210> 47
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 47
aataataac 9

<210> 48
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 48
tccaataaa 9

<210> 49
<211> 9
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 49

tctaattggg

9

<210> 50

<211> 17

<212> DNA

<213> Homo sapiens

<400> 50

tgcagtgtt cagccgc

17

<210> 51

<211> 18

<212> DNA

<213> Homo sapiens

<400> 51

tgcagtgtt cagccgct

18

<210> 52

<211> 18

<212> DNA

<213> Homo sapiens

<400> 52

ttgaagaagt cgtgtgtc

18

<210> 53

<211> 18

<212> DNA

<213> Homo sapiens

<400> 53

tgaagaagtc gtgtgtgt

18

<210> 54

<211> 17

<212> DNA

<213> Homo sapiens

<400> 54

tcgagctgaa gggcatc

17

<210> 55

<211> 18

<212> DNA

<213> Homo sapiens

<400> 55

tcgagctgaa gggcatcg

18

<210> 56

<211> 18

<212> DNA

<213> Homo sapiens

<400> 56
ttgtgccccca ggatgttg 18

<210> 57
<211> 18
<212> DNA
<213> Homo sapiens

<400> 57
tgtgccccag gatgttgc 18

<210> 58
<211> 137
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 58
tctagagaag acaagaacct gaccccagac caggtagtcg caatcgcgtc gaacattggg 60
ggaaagcaag ccctggaaac cgtgcaaagg ttgttgccgg tcctttgtca agaccacggc 120
cttaagagac cggatcc 137

<210> 59
<211> 137
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 59
tctagagaag acaagaacct gaccccagac caggtagtcg caatcgcgtc acatgacggg 60
ggaaagcaag ccctggaaac cgtgcaaagg ttgttgccgg tcctttgtca agaccacggc 120
cttaagagac cggatcc 137

<210> 60
<211> 137
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 60
tctagagaag acaagaacct gaccccagac caggtagtcg caatcgcgtc gaacaaaggg 60
ggaaagcaag ccctggaaac cgtgcaaagg ttgttgccgg tcctttgtca agaccacggc 120
cttaagagac cggatcc 137

<210> 61
<211> 137
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 61

tctagagaag acaagaacct gaccccagac caggtagtcg caatcgcgaa caataatggg 60

ggaaagcaag ccctggaaac cgtgcaaagg ttgttgccgg tcctttgtca agaccacggc 120

cttaagagac cggatcc 137

<210> 62

<211> 137

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 62

tctagagaag acaagaacct gaccccagac caggtagtcg caatcgcgtc aaacggaggg 60

ggaaagcaag ccctggaaac cgtgcaaagg ttgttgccgg tcctttgtca agaccacggc 120

cttaagagac cggatcc 137

<210> 63

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 63

tctagagaag acaacttaca ccggagcaag tcgtggccat tgcaagcaac atcgggtggca 60

aacaggctct tgagacgggt cagagacttc tcccagttct ctgtcaagcc cacgggctga 120

agagaccgga tcc 133

<210> 64

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 64

tctagagaag acaacttaca ccggagcaag tcgtggccat tgcattccac gacgggtggca 60

aacaggctct tgagacgggt cagagacttc tcccagttct ctgtcaagcc cacgggctga 120

agagaccgga tcc 133

<210> 65

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 65

tctagagaag acaacttaca ccggagcaag tcgtggccat tgcataaat aaaggtggca 60

aacaggctct tgagacgggt cagagacttc tcccagttct ctgtcaagcc cacgggctga 120

agagaccgga tcc 133

<210> 66

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 66

tctagagaag acaacttaca ccggagcaag tcgtggccat tgcaataat aacgggtggca 60

aacaggctct tgagacgggt cagagacttc tcccagttct ctgtcaagcc cacgggctga 120

agagaccgga tcc 133

<210> 67

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 67

tctagagaag acaacttaca ccggagcaag tcgtggccat tgcaagcaat gggggtggca 60

aacaggctct tgagacgggt cagagacttc tcccagttct ctgtcaagcc cacgggctga 120

agagaccgga tcc 133

<210> 68

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 68

tctagagaag acaactgact ccgatcaag ttgtagcgat tgcgtcgaac attggaggga 60

aacaagcatt ggagactgtc caacggctcc ttccgtgtt gtgtcaagcc cacggtttga 120

agagaccgga tcc 133

<210> 69

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 69

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtcgcat gacggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtttga 120

agagaccgga tcc 133

<210> 70

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 70

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtccaac aagggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtttga 120

agagaccgga tcc 133

<210> 71

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 71

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgaataac aatggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtttga 120

agagaccgga tcc 133

<210> 72

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 72

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtccaac ggtggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtttga 120

agagaccgga tcc 133

<210> 73

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 73

tctagagaag acaattgacg cctgcacaag tggtcgccat cgctccaat attggcggtg 60

agcaggcgct ggaaacagta cagcgctgc tgcctgtact gtgccaggat catggactga 120

agagaccgga tcc 133

<210> 74

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 74

tctagagaag acaattgacg cctgcacaag tggtcgccat cgccagccat gatggcggtg 60

agcaggcgct ggaaacagta cagcgctgc tgcctgtact gtgccaggat catggactga 120

agagaccgga tcc 133

<210> 75

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 75

tctagagaag acaattgacg cctgcacaag tggtcgccat cgccagcaat aagggcggtg 60

agcaggcgct ggaaacagta cagcgctgc tgcctgtact gtgccaggat catggactga 120

agagaccgga tcc 133

<210> 76

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 76

tctagagaag acaattgacg cctgcacaag tggtcgccat cgccaacaac aacggcggtg 60

agcaggcgct ggaaacagta cagcgctgc tgcctgtact gtgccaggat catggactga 120

agagaccgga tcc 133

<210> 77

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 77

tctagagaag acaattgacg cctgcacaag tggtcgccat cgcctcgaat ggcggcggtta 60

agcaggcgct ggaaacagta cagcgcctgc tgcctgtact gtgccaggat catggactga 120

agagaccgga tcc 133

<210> 78

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 78

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcgaac attgggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgcgggtcct ttgtcaagac cacggcctta 120

agagaccgga tcc 133

<210> 79

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 79

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcacat gacgggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgcgggtcct ttgtcaagac cacggcctta 120

agagaccgga tcc 133

<210> 80

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 80

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcgaac aaagggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgcgggtcct ttgtcaagac cacggcctta 120

agagaccgga tcc 133

<210> 81

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 81

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgaacaat aatgggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctta 120

agagaccgga tcc 133

<210> 82

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 82

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcaaac ggagggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctta 120

agagaccgga tcc 133

<210> 83

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 83

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtcgaac attggagggga 60

aacaagcatt ggagactgtc caacgggtcc ttcccgtgtt gtgtcaagcc cacgggtctga 120

agagaccgga tcc 133

<210> 84

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 84

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtcgcac gacggagggga 60

aacaagcatt ggagactgtc caacgggtcc ttcccgtgtt gtgtcaagcc cacgggtctga 120

agagaccgga tcc 133

<210> 85

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 85

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtccaac aagggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtctga 120

agagaccgga tcc 133

<210> 86

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 86

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgaataac aatggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtctga 120

agagaccgga tcc 133

<210> 87

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 87

tctagagaag acaactgact cccgatcaag ttgtagcgat tgcgtccaac ggtggaggga 60

aacaagcatt ggagactgtc caacggctcc ttcccggtgtt gtgtcaagcc cacggtctga 120

agagaccgga tcc 133

<210> 88

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 88

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcgaac attgggggaa 60

agcaagccct ggaaaccgtg caaaggttgt tgccggctct ttgtcaagac cacggcctga 120

agagaccgga tcc 133

<210> 89

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 89

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcacat gacgggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctga 120

agagaccgga tcc 133

<210> 90

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 90

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcgaac aaagggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctga 120

agagaccgga tcc 133

<210> 91

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 91

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgaacaat aatgggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctga 120

agagaccgga tcc 133

<210> 92

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 92

tctagagaag acaactgacc ccagaccagg tagtcgcaat cgcgtcaaac ggagggggaa 60

agcaagccct ggaaaccgtg caaagggttg tgccggtcct ttgtcaagac cacggcctga 120

agagaccgga tcc 133

<210> 93

<211> 36

<212> DNA

<213> Homo sapiens

<400> 93
 tcgccaccat ggtgagcaag ggcgaggagc tgttca 36

<210> 94
 <211> 36
 <212> DNA
 <213> Homo sapiens

<400> 94
 tggtgcccat cctggtcgag ctggacggcg acgtaa 36

<210> 95
 <211> 36
 <212> DNA
 <213> Homo sapiens

<400> 95
 tctgcaccac cggcaagctg cccgtgccct ggccca 36

<210> 96
 <211> 36
 <212> DNA
 <213> Homo sapiens

<400> 96
 tggagtacaa ctacaacagc cacaacgtct atatca 36

<210> 97
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 97
 ttcagcgtgt ccggcgaggg cgagggcgat gccaccta 38

<210> 98
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 98
 tgccacctac ggcaagctga ccctgaagtt catctgca 38

<210> 99
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 99
 tggcccaccc tcgtgaccac cctgacctac ggcgtgca 38

<210> 100
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 100
 ttcaagatcc gccacaacat cgaggacggc agcgtgca 38

<210> 101
 <211> 40
 <212> DNA
 <213> Homo sapiens

 <400> 101
 tagaggatcc accggtcgcc accatggtga gcaagggcga 40

 <210> 102
 <211> 40
 <212> DNA
 <213> Homo sapiens

 <400> 102
 tccggcgagg gcgagggcga tgccacctac ggcaagctga 40

 <210> 103
 <211> 40
 <212> DNA
 <213> Homo sapiens

 <400> 103
 tgacctacgg cgtgcagtgc ttcagccgct accccgacca 40

 <210> 104
 <211> 40
 <212> DNA
 <213> Homo sapiens

 <400> 104
 tccgccacaa catcgaggac ggcagcgtgc agctcgccga 40

 <210> 105
 <211> 42
 <212> DNA
 <213> Homo sapiens

 <400> 105
 tcctggtcga gctggacggc gacgtaaacg gccacaagtt ca 42

 <210> 106
 <211> 42
 <212> DNA
 <213> Homo sapiens

 <400> 106
 tcagccgcta ccccgaccac atgaagcagc acgacttctt ca 42

 <210> 107
 <211> 42
 <212> DNA
 <213> Homo sapiens

 <400> 107
 tcttcaagtc cgccatgccc gaaggctacg tccaggagcg ca 42

 <210> 108
 <211> 42
 <212> DNA
 <213> Homo sapiens

<400> 108
 tcaaggagga cggcaacatc ctggggcaca agctggagta ca 42

<210> 109
 <211> 42
 <212> DNA
 <213> Homo sapiens

<400> 109
 tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg ca 42

<210> 110
 <211> 44
 <212> DNA
 <213> Homo sapiens

<400> 110
 tccaccggtc gccaccatgg tgagcaaggg cgaggagctg ttca 44

<210> 111
 <211> 44
 <212> DNA
 <213> Homo sapiens

<400> 111
 ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaa 44

<210> 112
 <211> 44
 <212> DNA
 <213> Homo sapiens

<400> 112
 ttcagccgct accccgacca catgaagcag cacgacttct tcaa 44

<210> 113
 <211> 44
 <212> DNA
 <213> Homo sapiens

<400> 113
 tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaa 44

<210> 114
 <211> 46
 <212> DNA
 <213> Homo sapiens

<400> 114
 tggtgcccat cctggtcgag ctggacggcg acgtaaacgg ccacaa 46

<210> 115
 <211> 46
 <212> DNA
 <213> Homo sapiens

<400> 115
 tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacga 46

<210> 116

<211> 46
 <212> DNA
 <213> Homo sapiens

 <400> 116
 tgaaccgcat cgagctgaag ggcacgcgact tcaaggagga cggcaa 46

 <210> 117
 <211> 46
 <212> DNA
 <213> Homo sapiens

 <400> 117
 tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccga 46

 <210> 118
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 118
 tggtcaccgg ggtggtgccc atcctggctg agctggacgg cgacgtaa 48

 <210> 119
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 119
 tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttca 48

 <210> 120
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 120
 tgtccggcga gggcgagggc gatgccacct acggcaagct gaccctga 48

 <210> 121
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 121
 tcatctgcac caccggcaag ctgcccgtgc cctggccac cctcgtga 48

 <210> 122
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 122
 tctatatcat ggccgacaag cagaagaacg gcatcaaggt gaacttca 48

 <210> 123
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 123

tacggcgtgc agtgcttcag ccgctacccc gaccacatga agcagcacga 50

<210> 124
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 124
 taccgccgacc acatgaagca gcacgacttc ttcaagtccg ccatgcccgga 50

<210> 125
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 125
 tccgccatgc ccgaaggcta cgtccaggag cgcaccatct tcttcaagga 50

<210> 126
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 126
 ttcaaggagg acggcaacat cctggggcac aagctggagt acaactacaa 50

<210> 127
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 127
 tacaactaca acagccacaa cgtctatatc atggccgaca agcagaagaa 50

<210> 128
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 128
 tcaccggggt ggtgcccac cctggtcgagc tggacggcga cgtaaaccggc ca 52

<210> 129
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 129
 taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc ta 52

<210> 130
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 130
 tgcagtgcct cagccgctac cccgaccaca tgaagcagca cgacttcttc aa 52

<210> 131
 <211> 54

<212> DNA
 <213> Homo sapiens

 <400> 131
 tcagcgtgtc cggcgagggc gagggcgatg ccacctacgg caagctgacc ctga 54

 <210> 132
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 132
 tgaagttcat ctgcaccacc ggcaagctgc ccgtgccctg gcccaccctc gtga 54

 <210> 133
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 133
 tcgtgaccac cctgacctac ggcgtgcagt gcttcagccg ctaccccgac caca 54

 <210> 134
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 134
 tcatggccga caagcagaag aacggcatca aggtgaactt caagatccgc caca 54

 <210> 135
 <211> 56
 <212> DNA
 <213> Homo sapiens

 <400> 135
 ttcaccgggg tggtgcccac cctggctcag ctggacggcg acgtaaacgg ccacaa 56

 <210> 136
 <211> 56
 <212> DNA
 <213> Homo sapiens

 <400> 136
 tgcaccaccg gcaagctgcc cgtgccctgg cccaccctcg tgaccaccct gaccta 56

 <210> 137
 <211> 56
 <212> DNA
 <213> Homo sapiens

 <400> 137
 ttcaagtccg ccatgcccga aggctacgtc caggagcgca ccatttttt caagga 56

 <210> 138
 <211> 56
 <212> DNA
 <213> Homo sapiens

 <400> 138
 tatatcatgg ccgacaagca gaagaacggc atcaaggtga acttcaagat ccgcca 56

<210> 139
 <211> 58
 <212> DNA
 <213> Homo sapiens

 <400> 139
 ttcac tctgca ccaccggcaa gctgcccgtg ccctggccca ccctcgtgac caccctga 58

 <210> 140
 <211> 58
 <212> DNA
 <213> Homo sapiens

 <400> 140
 tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc gacttcaa 58

 <210> 141
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 141
 tacctattat tactttatgg ggcagcagcc tggaaaagta cttggggacc aa 52

 <210> 142
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 142
 tgtgtcttgg gatgagtggg tcagtgttct ggtgctcaca ggatggctgg ca 52

 <210> 143
 <211> 51
 <212> DNA
 <213> Homo sapiens

 <400> 143
 tcctgtggct cctgccgctg ctgctttcca cggcagctgt gggctccggg a 51

 <210> 144
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 144
 tatgtacgcc tccctgggct cgggtccggt cgcccctttg cccgcttctg ta 52

 <210> 145
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 145
 tgaattggga tgctgttttt aggtattcta ttcaaattta ttttactgtc tttta 54

 <210> 146
 <211> 52
 <212> DNA

<213> Homo sapiens
 <400> 146
 tccctcacca tgagtagcgc tatgttggtg acttgccctcc cggaccccag ca 52
 <210> 147
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 147
 tgtgcgatct ccaagcactg aggggcagaa actcccggat cgggcgctgc ca 52
 <210> 148
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 148
 ttttcaagtg aagacaaaat ggcctcgccg gctgacagct gtatccagtt ca 52
 <210> 149
 <211> 53
 <212> DNA
 <213> Homo sapiens
 <400> 149
 tacaattgaa caatgcctca gctatacatt tacatcagat tattgggagc cta 53
 <210> 150
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 150
 tccgaagctg acagatgggt attctttgac ggggggtagg ggcggaacct ga 52
 <210> 151
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 151
 ttagacttag gtaagtaatg caatatggtg gactggggag aactacaaac ta 52
 <210> 152
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 152
 tctgcaataa aaaatggcct ccaacaaaac tacattggta agttaatgaa aa 52
 <210> 153
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 153
 tggagctttc agcggtgggg gagcgggtgt tcgcggccga agccctcctg aa 52

<210> 154
 <211> 55
 <212> DNA
 <213> Homo sapiens

 <400> 154
 tggaacacca gctcctgtgc tgcgaagtgg aaaccatccg ccgcgcgtac cccga 55

 <210> 155
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 155
 tgcttagcgt cctgcgacag tacaacatcc agaagaagga gattgtggtg aa 52

 <210> 156
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 156
 tgctgcaggt accccggatc ccctgacttg cgagggacgc attcgggccg ca 52

 <210> 157
 <211> 49
 <212> DNA
 <213> Homo sapiens

 <400> 157
 tcccttgatc tgagaatggc tacctctcga tatgagccag tggctgaaa 49

 <210> 158
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 158
 tggcgtcggg cctgggctcc ccgtccccct gctcggcggg cagtgaggag ga 52

 <210> 159
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 159
 tgtgttgga gaagatggca gatccaggaa tgatgagtct ttttggcgag ga 52

 <210> 160
 <211> 51
 <212> DNA
 <213> Homo sapiens

 <400> 160
 tccagcgtgg acaatggcta ctcaaggttt gtgtcattaa atctttagtt a 51

 <210> 161
 <211> 54
 <212> DNA
 <213> Homo sapiens

<400> 161
 taatatcaca atgagttcag gcttatggag ccaagaaaaa gtcacttcac ccta 54

<210> 162
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 162
 tcacacggag gacgcgatgg ctccaagaa acgcccagaa acccagaaga 50

<210> 163
 <211> 49
 <212> DNA
 <213> Homo sapiens

<400> 163
 tccggccggc gccatgaagt gagaaggggg ctgggggtcg cgctcgcta 49

<210> 164
 <211> 55
 <212> DNA
 <213> Homo sapiens

<400> 164
 tccgggatcg ccatgggaac tcaatagaaa atcctcatct tctcactttg tttca 55

<210> 165
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 165
 tggcgtccac gggtagtat ggtggaactg cggtcgcgcc ggcggtagcc gga 53

<210> 166
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 166
 tgaccaggc aggacacatg caggccaaaa aacgctatct catcctgctc tca 53

<210> 167
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 167
 ttctcccag ggggatgtcc tgcgcctcag ggtccggtgg tggcctgcgg ca 52

<210> 168
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 168
 tgcttttaga ataatcatgg gccagactgg gaagaaatct gagaagggac ca 52

<210> 169
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 169
 taggcgccaa ggccatgtcc gactcgtggg tcccgaactc cgcctcgggc ca 52

 <210> 170
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 170
 tgaagggaca tcaccttttc gctttttcca agatggctca agattcagta ga 52

 <210> 171
 <211> 53
 <212> DNA
 <213> Homo sapiens

 <400> 171
 tgccccggca tggcgacacc ggacgcgggg ctccctgggg ctgagggcgt gga 53

 <210> 172
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 172
 ttcgcgacc tcatggaatc ctttctgcag cacctggatc gcttttccga 50

 <210> 173
 <211> 50
 <212> DNA
 <213> Homo sapiens

 <400> 173
 tcggccacca tgtcccgcca gaccacctct gtgggctcca gctgcctgga 50

 <210> 174
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 174
 tccccagaac agcactatgg gcttctcttc cgagctgtgc agccccagg gcca 54

 <210> 175
 <211> 51
 <212> DNA
 <213> Homo sapiens

 <400> 175
 tctgctcccc accgaggacc tctgcatgca ggcatgaatc ccaggagcct a 51

 <210> 176
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 176
 tgtaccgagc acttcggctc ctcgcgcgct cgcgtcccct cgtgcgggct cca 53

<210> 177
 <211> 54
 <212> DNA
 <213> Homo sapiens

<400> 177
 tctccaaggc accatgaatg ccatcgtggc tctctgccac ttctgcgagc tcca 54

<210> 178
 <211> 56
 <212> DNA
 <213> Homo sapiens

<400> 178
 tccggaggcc atgccggcgt tggcgcgcga cggcggccag ctgccgctgc tcggta 56

<210> 179
 <211> 55
 <212> DNA
 <213> Homo sapiens

<400> 179
 tgcagcgggg cgccgcgctg tgcctgcgac tgtggctctg cctgggactc ctgga 55

<210> 180
 <211> 49
 <212> DNA
 <213> Homo sapiens

<400> 180
 tcacatggc cgaggcgcct caggtggtgg agatcgaccc ggacttcga 49

<210> 181
 <211> 54
 <212> DNA
 <213> Homo sapiens

<400> 181
 tctccgctcg aagtggagct ggacccggag ttcgagcccc agagccgtcc gcga 54

<210> 182
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 182
 tcctctgaga cgccatgttc aactcgatga cccaccacc aatcagtagc ta 52

<210> 183
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 183
 tggcgcagac gcagggcacc cggaggaaag tctgttacta ctacgacggt ga 52

<210> 184

<211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 184
 tgcgctcacc tccctgcggc ctcctgaggt ggtttggtgg cccctcctc gcga 54

 <210> 185
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 185
 tcctcaacta tgacctcaac cggccaggat tccaccacaa ccaggcagcg aa 52

 <210> 186
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 186
 tgagcgcacg cggtgagggc gcggggcagc cgtccacttc agcccaggga ca 52

 <210> 187
 <211> 53
 <212> DNA
 <213> Homo sapiens

 <400> 187
 tccgtgctcc tccacccccg ctggatcgag cccaccgtca tgtttctcta cga 53

 <210> 188
 <211> 49
 <212> DNA
 <213> Homo sapiens

 <400> 188
 tgggcacggt gatggccacc actggggccc tgggcaacta ctacgtgga 49

 <210> 189
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 189
 tccagcagat catgtcatga cgacttcgct gtcctgcat ccacgctggc cgga 54

 <210> 190
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 190
 ttgacgagtg cggccagagc gcagccagca tgtacctgcc gggctgcgcc ta 52

 <210> 191
 <211> 53
 <212> DNA
 <213> Homo sapiens

 <400> 191

tgcgggcaga cggcgggggc gccgggtggcg ccccggcctc ttcctcctcc tca 53

<210> 192
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 192
 tctgaaaaag actctgcatg ggaatggcct gccttacgat gacagaaatg ga 52

<210> 193
 <211> 55
 <212> DNA
 <213> Homo sapiens

<400> 193
 taccgcgatg agaggcgctc gcggcgccctg ggattttctc tgcgttctgc tccta 55

<210> 194
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 194
 tgaaaatgac tgaatataaa cttgtggtag ttggagctgg tggcgtaggc aa 52

<210> 195
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 195
 taggtcccc ggcgccaggc caccggccg tcagcagcat gcagggtaag ga 52

<210> 196
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 196
 tccaagcgcg aaaaccccg atggtgagga gcaggtactg gcccggcagc ga 52

<210> 197
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 197
 ttattattac atggctttgc cttactgagg cttcatcttg tcctctggtc ca 52

<210> 198
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 198
 tctggcgcca aaatgtcgtt cgtggcaggg gttattcggc ggctggacga ga 52

<210> 199
 <211> 52

<212> DNA
 <213> Homo sapiens
 <400> 199
 tgaggaggtt tcgacatggc ggtgcagccg aaggagacgc tgcagttgga ga 52

<210> 200
 <211> 54
 <212> DNA
 <213> Homo sapiens
 <400> 200
 tcactgtcgg cgcccatgac accgctcgtc tccgcctga gtcgtctgtg ggta 54

<210> 201
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 201
 tgcttagacg ctggatTTTT ttcgggtagt ggaaaaccag gtaagcaccg aa 52

<210> 202
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 202
 tccgcaggg agcggacatg gactacgact cgtaccagca ctatttctac ga 52

<210> 203
 <211> 50
 <212> DNA
 <213> Homo sapiens
 <400> 203
 tgccgagctg ctccacgtcc accatgccgg gcatgatctg caagaacca 50

<210> 204
 <211> 48
 <212> DNA
 <213> Homo sapiens
 <400> 204
 tgaggagccg gaccgatgtg gaaactgctg cccgccgagg gcccgga 48

<210> 205
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 205
 tctttactga taatgtcaag ttcattgtac cttccaacc aaggagcatt ca 52

<210> 206
 <211> 48
 <212> DNA
 <213> Homo sapiens
 <400> 206
 tggagggcca ctgagccccg ctaccgccc cacagccttt cctacca 48

<210> 207
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 207
 tcggcgcatg aaggaggtac tcctcatttt cgttctctct ctctgtgccc ca 52

 <210> 208
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 208
 ttgcgctcgg ggcggccatg tcggccggcg aggtcgagcg cctagtgtcg ga 52

 <210> 209
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 209
 tctgcaggac accatgcggc ttccgggtgc gatgccagct ctggccctca aa 52

 <210> 210
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 210
 tgagtactcc gcctctaccc cggctgaagc ccgccccgc cgccacctat ta 52

 <210> 211
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 211
 tcgggtgttg catccatgga gcgagctgag agctcgaggt gagcggggct cgca 54

 <210> 212
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 212
 tggaactgct taatagaaac aggcttgtaa ttgtgagtcc gcgctgca 48

 <210> 213
 <211> 51
 <212> DNA
 <213> Homo sapiens

 <400> 213
 tcccagacat gacagccatc atcaaagaga tcgttagcag aaacaaaagg a 51

 <210> 214
 <211> 53
 <212> DNA

<213> Homo sapiens
 <400> 214
 tggcatggcc agcaacagca gtcctgtccc gacacctggg ggcgggcacc tca 53
 <210> 215
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 215
 tgctgggtga gaagggctgt ggctgcgttt tagagaagcg ttgggtactg ga 52
 <210> 216
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 216
 tgcgggacgt gcgggagcgg ctgcaggcgt gggagcgcgc gttccgacgg ca 52
 <210> 217
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 217
 tcagaataca gttatggcca cccaggtaat ggggcagtct tctggaggag ga 52
 <210> 218
 <211> 53
 <212> DNA
 <213> Homo sapiens
 <400> 218
 tgagtctgc cggccgccgg ctcccgcagg ggccagggcg aagttggcgc cga 53
 <210> 219
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 219
 ttctttattt ccagcaatgt ctcaggctgt gcagacaaac ggaactcaac ca 52
 <210> 220
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 220
 ttcaggagga agcgatggct tcagacagca tatttgagtc atttccttcg ta 52
 <210> 221
 <211> 52
 <212> DNA
 <213> Homo sapiens
 <400> 221
 ttccttgag gcgccggttg ccggccacaa cccttggcgg agcctgcctg ca 52

<210> 222
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 222
 tgttgctgag gtgacttcag tgggactggg agttggtgcc tgcggccctc cgga 54

 <210> 223
 <211> 53
 <212> DNA
 <213> Homo sapiens

 <400> 223
 tcaggaacga gatggcgggt ctctggaggc tgagtgccgt ttgcggtgcc cta 53

 <210> 224
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 224
 tgcagaggac aaaagcatgt cttcccttcc tgggtgcatt ggtttgatg ca 52

 <210> 225
 <211> 54
 <212> DNA
 <213> Homo sapiens

 <400> 225
 ttacgcggcg gggctgtcgc cgtacgcgga caagggcaag tgcggcctcc cgga 54

 <210> 226
 <211> 52
 <212> DNA
 <213> Homo sapiens

 <400> 226
 tttgtaaga acatgtcgtc catcttgcca ttcacgccgc cagttgtgaa ga 52

 <210> 227
 <211> 53
 <212> DNA
 <213> Homo sapiens

 <400> 227
 tggtgacggc ggcaacatgt ctgtggcttt cgcggcccg aggcagcgag gca 53

 <210> 228
 <211> 48
 <212> DNA
 <213> Homo sapiens

 <400> 228
 tggcgctca gaagcacggc ggtgggggag ggggcggctc ggggccca 48

 <210> 229
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 229
 tcatgtctca tgcggccgaa ccagctcggg atggcgtaga ggccagcgcg ga 52

<210> 230
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 230
 tcgggggctg ctcagggggc tgtggccgct gcacatcgct ctgtggacgc gta 53

<210> 231
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 231
 ttccgcccgc ccaggatgga ggcgcccgc agcgcgcaga ccccgcaccc gca 53

<210> 232
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 232
 ttgccgtccc aagcaatgga tgatttgatg ctgtccccgg acgatattga aca 53

<210> 233
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 233
 tcctggtcca ccatggccaa accaacaagc aaagattcag gcttgaagga ga 52

<210> 234
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 234
 tctggatcgc ggagggaatg ccccgaggag cggagaactg ggacgaggcc ga 52

<210> 235
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 235
 tgggccagag atggcgggcg ccgacggggc tttgccggag gcggcggtt ta 52

<210> 236
 <211> 55
 <212> DNA
 <213> Homo sapiens

<400> 236
 tgcccagaca agcaacatgg ctcggaaacg cgcggccggc ggggagccgc gggga 55

<210> 237
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 primer

 <220>
 <223> 5'-biotin

 <400> 237
 tctagagaag acaagaacct gacc 24

<210> 238
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 primer

 <400> 238
 ggatccggtc tcttaaggcc gtgg 24

<210> 239
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 primer

 <400> 239
 gacgtgggct gtcaaatacc aagatatg 28

<210> 240
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 primer

 <400> 240
 tctcctccag ttcacttttg actagttggg 30

<210> 241
 <211> 20
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 primer

 <400> 241
 agtaacagcg gtagaggcag 20

<210> 242
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic primer

<400> 242
 attgggctac gatggactcc 20

<210> 243
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic primer

<400> 243
 ttaattcaat atattcatga ggcac 25

<210> 244
 <211> 5656
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<220>
 <221> modified_base
 <222> (901)..(910)
 <223> a, c, t, g, unknown or other

<400> 244
 gacggatcgg gagatctccc gatcccctat ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
 ttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240
 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300
 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggtggac tatttacggg aaactgccc cttggcagta catcaagtgt 480
 atcatatgcc aagtacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600
 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660
 actcacgggg atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc 720
 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggagc 780

gtaggcgtgt	acggtgggag	gtctatataa	gcagagctct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagacccaa	gctggctagc	900
nnnnnnnnnn	ggatcccaac	tagtcaaaag	tgaactggag	gagaagaaat	ctgaacttcg	960
tcataaattg	aaatatgtgc	ctcatgaata	tattgaatta	attgaaattg	ccagaaattc	1020
cactcaggat	agaattcttg	aatgaaggt	aatggaattt	tttatgaaag	tttatggata	1080
tagaggtaaa	catttgggtg	gatcaaggaa	accggacgga	gcaatttata	ctgtcggatc	1140
tcctattgat	tacggtgtga	tcgtggatac	taaagcttat	agcggagggt	ataatctgcc	1200
aattggccaa	gcagatgaaa	tgcaacgata	tgtcgaagaa	aatcaaacac	gaaacaaaca	1260
tatcaaccct	aatgaatggt	ggaaagtcta	tccatcttct	gtaacggaat	ttaagttttt	1320
atttgtgagt	ggtcacttta	aaggaaacta	caaagctcag	cttacacgat	taaatcatat	1380
cactaattgt	aatggagctg	ttcttagtgt	agaagagctt	ttaattgggtg	gagaaatgat	1440
taaagccggc	acattaacct	tagaggaagt	cagacggaaa	tttaataacg	gcgagataaa	1500
cttttaaggg	cccttcgaag	gtaagcctat	ccctaaccct	ctcctcggtc	tcgattctac	1560
gcgtaccggt	catcatcacc	atcaccattg	agtttaaacc	cgctgatcag	cctcgactgt	1620
gccttctagt	tgccagccat	ctgtttgtttg	cccctcccc	gtgccttcct	tgaccctgga	1680
aggtgccact	cccactgtcc	tttcctaata	aaatgaggaa	attgcatcgc	attgtctgag	1740
taggtgtcat	tctattctgg	ggggtggggt	ggggcaggac	agcaaggggg	aggattggga	1800
agacaatagc	aggcatgctg	gggatgcggt	gggctctatg	gcttctgagg	cggaaagaac	1860
cagctggggc	tctagggggg	atccccacgc	gccctgtagc	ggcgcattaa	gcgcggcggg	1920
tgtggtgggt	acgcgcagcg	tgaccgctac	acttgccagc	gccctagcgc	ccgctccttt	1980
cgctttcttc	ccttcctttc	tcgccacgtt	cgccggcttt	ccccgtcaag	ctctaaatcg	2040
gggcatccct	ttaggggttc	gatttagtgc	tttacggcac	ctcgacccca	aaaaacttga	2100
ttagggtgat	ggttcacgta	gtgggccatc	gccctgatag	acggtttttc	gccctttgac	2160
gttgaggtcc	acgttcttta	atagtggact	cttgttccaa	actggaacaa	cactcaaccc	2220
tatctcggtc	tattcttttg	atttataagg	gattttgggg	atttcggcct	attggttaaa	2280
aaatgagctg	atttaacaaa	aatttaacgc	gaattaattc	tgtggaatgt	gtgtcagtta	2340
gggtgtggaa	agtccccagg	ctccccaggc	aggcagaagt	atgcaaagca	tgcattctcaa	2400
ttagtcagca	accaggtgtg	gaaagtcccc	aggctcccca	gcaggcagaa	gtatgcaaag	2460
catgcatctc	aattagtcag	caaccatagt	cccgccctta	actccgcca	tcccgcccct	2520
aactccgccc	agttccgccc	attctccgcc	ccatggctga	ctaatttttt	ttatttatgc	2580
agaggccgag	gccgcctctg	cctctgagct	attccagaag	tagtgaggag	gcttttttgg	2640
aggcctaggc	ttttgcaaaa	agctcccggg	agcttgata	tccatttttcg	gatctgatca	2700
gcacgtgttg	acaattaatc	atcggcatag	tatatcgga	tagtataata	cgacaagggtg	2760
aggaactaaa	ccatggccaa	gcctttgtct	caagaagaat	ccaccctcat	tgaaagagca	2820

acggctacaa	tcaacagcat	ccccatctct	gaagactaca	gcgtcgccag	cgcagctctc	2880
tctagcgacg	gccgcatctt	cactggtgtc	aatgtatatc	attttactgg	gggaccttgt	2940
gcagaactcg	tggtgctggg	cactgctgct	gctgcggcag	ctggcaacct	gacttgtatc	3000
gtcgcgatcg	gaaatgagaa	caggggcata	ttgagcccct	gcggacggtg	tcgacaggtg	3060
cttctcgatc	tgcatacttg	gatcaaagcg	atagtgaagg	acagtgatgg	acagccgacg	3120
gcagttggga	ttcgtgaatt	gctgccctct	ggttatgtgt	gggaggggcta	agcacttcgt	3180
ggccgaggag	caggactgac	acgtgctacg	agatttcgat	tccaccgccg	ccttctatga	3240
aaggttgggc	ttcggaatcg	ttttccggga	cgccggctgg	atgatcctcc	agcgcgggga	3300
tctcatgctg	gagttcttcg	cccaccccaa	cttgttttatt	gcagcttata	atggttacaa	3360
ataaagcaat	agcatcacia	atttcacaaa	taaagcattt	ttttcactgc	attctagttg	3420
tggtttgtcc	aaatcatca	atgtatctta	tcatgtctgt	ataccgtcga	cctctagcta	3480
gagcttggcg	taatcatggt	catagctgtt	tcctgtgtga	aattgtttatc	cgctcacaat	3540
tccacacaa	atacgagccg	gaagcataaa	gtgtaaagcc	tggggtgcct	aatgagtgag	3600
ctaactcaca	ttaattgcgt	tgcgtcact	gcccgttttc	cagtcgggaa	acctgtcgtg	3660
ccagctgcat	taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttgggcgctc	3720
ttccgcttcc	tcgctcactg	actcgtgcg	ctcggtcgtt	cggctgcggc	gagcggtatc	3780
agctcactca	aaggcggtaa	tacggttatc	cacagaatca	ggggataacg	caggaaagaa	3840
catgtgagca	aaaggccagc	aaaaggccag	gaaccgtaaa	aaggccgcgt	tgctggcggtt	3900
tttccatagg	ctccgcccc	ctgacgagca	tcacaaaaat	cgacgctcaa	gtcagaggtg	3960
gcgaaacccg	acaggactat	aaagatacca	ggcgtttccc	cctggaagct	ccctcgtgcg	4020
ctctcctgtt	ccgacctgc	cgcttaccgg	atacctgtcc	gcctttctcc	cttcgggaag	4080
cgtggcgctt	tctcaatgct	cacgctgtag	gtatctcagt	tcggtgtagg	tcgttcgctc	4140
caagctgggc	tgtgtgcacg	aacccccctg	tcagcccagc	cgctgcgcct	tatccggtaa	4200
ctatcgtctt	gagtccaacc	cggtaaagaca	cgacttatcg	ccactggcag	cagccactgg	4260
taacaggatt	agcagagcga	ggtatgtagg	cggtgctaca	gagttcttga	agtgggtggcc	4320
taactacggc	tacactagaa	ggacagtatt	tggtatctgc	gctctgctga	agccagttac	4380
cttcggaaaa	agagttggta	gctcttgatc	cggcaaacaa	accaccgctg	gtagcgggtg	4440
tttttttgtt	tgcaagcagc	agattacgcg	cagaaaaaaa	ggatctcaag	aagatccttt	4500
gatcttttct	acggggctctg	acgctcagt	gaacgaaaac	tcacgttaag	ggattttggt	4560
catgagatta	tcaaaaagga	tcttcaccta	gatcctttta	aattaaaaat	gaagttttaa	4620
atcaatctaa	agtatatatg	agtaaacttg	gtctgacagt	taccaatgct	taatcagtga	4680
ggcacctatc	tcagcgatct	gtctatttcg	ttcatccata	gttgccctgac	tccccgtcgt	4740
gtagataact	acgatacggg	agggcttacc	atctggcccc	agtgctgcaa	tgataccgcg	4800
agaccacgc	tcaccggctc	cagatttatc	agcaataaac	cagccagccg	gaagggccga	4860

29539_0036W01_SL2953936W.txt

gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga	4920
agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca ttgctacagg	4980
catcgtgggtg tcacgctcgt cgtttggtat ggcttcattc agtccgggtt cccaacgatc	5040
aaggcgagtt acatgatccc ccatgtttgtg caaaaaagcg gttagctcct tcggtcctcc	5100
gatcgtttgtc agaagtaagt tggccgcagt gttatcactc atgggttatgg cagcactgca	5160
taatttcttt actgtcatgc catccgtaag atgcttttct gtgactgggtg agtactcaac	5220
caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg	5280
ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa aacgtttcttc	5340
ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcg	5400
tgcaccaaac tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac	5460
aggaaggcaa aatgccgcaa aaaagggaaat aagggcgaca cggaaatgtt gaatactcat	5520
actcttcctt tttcaatatt attgaagcat ttatcagggg tattgtctca tgagcggata	5580
catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa	5640
agtgccacct gacgtc	5656

<210> 245

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 245

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttact	240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtagtcgc aatcgcgtcg aacattgggg gaaagcaagc cctggaaacc	600
gtgcaaagggt tgttgccggg cttttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcaa gcaatggggg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat	780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac	840

29539_0036W01_SL2953936W.txt

```

ggtttgacgc ctgcacaagt ggtcgccatc gcctcgaatg gcggcggtaa gcaggcgctg      900
gaaacagtac agcgctgct gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtcgcaa tcgcgtcgaa cattggggga aagcaagccc tggaaaccgt gcaaaggttg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcatcc     1080
cacgacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gccacggggc tgactcccg a tcaagttgta gcgattgcga ataacaatgg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgttggtgc aagcccacgg tttgacgcct     1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtaagc aggcgctgga aacagtacag     1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc     1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa tgggggtggc     1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaacaagc attggagact     1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc     1680
gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct     1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcacatgac     1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac     1860
ggccttacac cggagcaagt cgtggccatt gcaagcaatg ggggtggcaa acaggctctt     1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa     1980
gtttagcga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc     2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccaac     2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag     2160
gatcatggac tgacacccga acaggtggtc gccattgcta ataataacgg aggacggcca     2220
gccttggagt ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg     2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa     2340
aagggtctgc ctcatgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga     2400
acttcccatc gagtgcgagg atcc                                     2424

```

<210> 246

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 246

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

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

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
580 585 590

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
610 615 620

Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 247
<211> 2424
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 247

gctagcacca	tggactacaa	agaccatgac	ggtgattata	aagatcatga	catcgattac	60
aaggatgacg	atgacaagat	ggccccaag	aagaagagga	aggtgggcat	tcaccgctgg	120
gtacctatgg	tggacttgag	gacactcggt	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaagggtca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggcttcact	240
catgcgcata	ttgtcgcgt	ttcacagcac	cctgcgccgc	ttgggacggt	ggctgtcaaa	300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcggg	360
aaacagtggg	cgggagcgcg	agcacttgag	gcgctgctga	ctgtggcggg	tgagcttagg	420
gggctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcaagag	agggggagta	480
acagcggtag	aggcagtgc	cgctggcgc	aatgcgctca	ccggggcccc	cttgaacctg	540
acccagacc	aggtagtcgc	aatcgctcg	aacattgggg	gaaagcaagc	cctggaaacc	600
gtgcaaaggt	tggtgccggt	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcat	cccacgacgg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagccacagg	gctgactccc	gatcaagttg	tagcgattgc	gtcgaacatt	780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgt	tcaagccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gccaacaaca	acggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgctcgaa	cattggggga	aagcaagccc	tggaaaccgt	gcaaagggtg	1020
ttgccggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aacatcggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gccacgggc	tgactccga	tcaagttgta	gcgattgcga	ataacaatgg	agggaaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgttggtgc	aagccacagg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	cagccatgat	ggcggttaagc	aggcgctgga	aacagtacag	1320
cgcctgctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccaggt	agtcgcaatc	1380
gcgaacaata	atgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttggt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaaataa	taacgggtggc	1500
aaacaggctc	ttgagacggt	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgaat	aacaatggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccg	gttgtgtcaa	gccacgggtt	tgacgcctgc	acaagtggtc	1680
gcatcgcca	gcatgatgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcgaacatt	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac	1860

```

ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgtcgaa cattggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgccaac 2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacacccga acaggtggtc gccattgcta ataataacgg aggacggcca 2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggctctgc ctcatgtctc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gagtcgcggg atcc 2424

```

<210> 248

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 248

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 249

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 249

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcg	120
gtacctatgg tggacttgag gacactcggt tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcgg	360
aaacagtggg cgggagcgc agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagtcga caccgggcag ctgctgaaga tcgcaagag agggggagta	480
acagcggtag aggcagtgc cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtatgcg aatcgcgtca catgacgggg gaaagcaagc cctggaacc	600
gtgcaaaggt tggtgccgt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcat cccacgacg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat	780

```

ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgttggtg tcaagccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gcctccaata ttggcggtaa gcaggcgctg      900
gaaacagtac agcgctgtct gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtgcgaa tcgctgcgaa cattggggga aagcaagccc tggaaaccgt gcaaagggtg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat     1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gccacgggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgttggtgc aagcccacgg tttagcgcct     1260
gcacaagtgg tcgccatcgc ctcgaatggc ggcggtaagc aggcgctgga aacagtacag     1320
cgctgtctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggc agtcgcaatc     1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaagggttgtt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa catcggtggc     1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaaacaagc attggagact     1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc     1680
gccatcgctt ccaatattgg cggtaaagcag gcgctggaaa cagtacagcg cctgctgcct     1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gaacaataat     1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac     1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt     1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa     1980
gttgtagcga ttgcgtccaa cgggtggagg aaacaagcat tggagactgt ccaacggctc     2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccaac     2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag     2160
gatcatggac tgacaccca acaggtggtc gccattgcta ataataacgg aggacggcca     2220
gccttgagat ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg     2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa     2340
aagggtctgc ctcatgtctc cgcattgatc aaaagaacca accggcggat tcccgagaga     2400
acttcccatc gagtgcgagg atcc                                     2424

```

<210> 250

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 250

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His

1 5 10 15
 Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
 20 25 30
 Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
 35 40 45
 Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 50 55 60
 Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80
 His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95
 Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110
 His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125
 Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 Page 92

275

280

285

Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 Page 93

545 550 555 560
 Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 690 695 700
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720
 Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
 725 730 735
 Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750
 Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765
 Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780
 His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800
 Thr Ser His Arg Val Ala Gly Ser
 805

<210> 251

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 251

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcaagag agggggagta	480
acagcggtag aggcatgca cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
accccagacc aggtagtcgc aatcgcgtca catgacgggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcaa gcaacatcgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat	780
ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gccaacaaca acggcggtaa gcaggcgctg	900
gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac cccagaccag	960
gtagtgcgaa tcgctgcaaa cggaggggga aagcaagccc tggaaaccgt gcaaaggttg	1020
ttgccgggtc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc	1080
aatgggggtg gcaaacaggc tcttgagacg gttagagac ttctcccagt tctctgtcaa	1140
gcccacgggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagcccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtgaagc aggcgctgga aacagtacag	1320
cgctgtgtgc ctgtactgtg ccaggatcat ggactgacct cagaccaggg agtcgcaatc	1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcatcca cgacgggtggc	1500
aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaaacaagc attggagact	1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtgggc	1680
gcatcgcca gccatgatgg cggtgaagcag gcgctggaaa cagtacagcg cctgctgcct	1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcacatgac	1800

```

gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaagcaatg ggggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgtcgaa cattggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgccagc 2100
catgatggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacacccga acaggtggtc gccattgctt cccacgacgg aggacggcca 2220
gccttgagat ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggtctgc ctcatgctcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gagtgcggg atcc 2424

```

<210> 252

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 252

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 253

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 253

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
accccagacc aggtatgcgc aatcgcgtca aacggagggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcaa gcaacatcgg tggcaaacag gctcttgaga cggttcagag acttctccca	720

29539_0036W01_SL2953936W.txt

gttctctgtc aagccacg gctgactccc gatcaagttg tagcgattgc gaataacaat	780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgttgtg tcaagccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gcctccaata ttggcggtaa gcaggcgctg	900
gaaacagtac agcgctgtct gcctgtactg tgccaggatc atggactgac cccagaccag	960
gtagtgcgaa tcgctgcaca tgacggggga aagcaagccc tggaaccgt gcaaaggttg	1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc	1080
aatgggggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa	1140
gccacgggc tgactcccga tcaagttgta gcgattgcgt ccaacgggtg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc ctccaatatt ggcggtgaagc aggcgctgga aacagtacag	1320
cgctgtgtc ctgtactgtg ccaggatcat ggactgacc cagaccagg agtcgcaatc	1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaaataa taacggtggc	1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560
actccgatc aagttgtagc gattgcgtcc aacggtggag ggaacaagc attggagact	1620
gtccaacggc tccttcccgt gttgtgtcaa gccacggtt tgacgcctgc acaagtggtc	1680
gccatcgcct ccaatattgg cggtaagcag gcgctggaaa cagtacagcg cctgtgcct	1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcgaacatt	1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac	1860
ggccttacac cggagcaagt cgtggccatt gcaaataata acggtggcaa acaggctctt	1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tccgatcaa	1980
gttgtagcga ttgcgtccaa cgggtgaggg aaacaagcat tggagactgt ccaacggctc	2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgcctcc	2100
aatattggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag	2160
gatcatggac tgacacccga acagggtggtc gccattgctt ctaacatcgg aggacggcca	2220
gccttgagat ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg	2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa	2340
aagggtctgc ctcatgtctc cgattgatc aaaagaacca accggcggat tcccagaga	2400
acttcccatc gactcgcggt atcc	2424

<210> 254

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 254

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

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

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 690 695 700
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720
 Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile
 725 730 735
 Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750
 Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765
 Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780
 His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800
 Thr Ser His Arg Val Ala Gly Ser
 805

<210> 255
 <211> 2424
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 255
 gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60
 aaggatgacg atgacaagat ggccccaag aagaagagga aggtgggcat tcaccgcggg 120
 gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag 180
 cctaagggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact 240
 catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa 300
 taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg 360
 aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420
 gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480
 acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg 540
 accccagacc aggtagtcgc aatcgcgtcg aacattgggg gaaagcaagc cctggaaacc 600
 gtgcaaagggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg 660
 gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
 gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtccaacggg 780
 ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgttgtg tcaagcccac 840
 ggtttgacgc ctgcacaagt ggtcgccatc gcctcgaatg gcggcggtaa gcaggcgctg 900
 gaaacagtac agcgctgct gcctgtactg tgccaggatc atggactgac cccagaccag 960
 gtagtcgcaa tcgctgcaaa cggaggggga aagcaagccc tggaaaccgt gcaaagggtg 1020
 ttgccgggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat 1080
 aataacgggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
 gccacgggc tgactcccga tcaagttgta gcgattgcgt ccaacgggtg agggaaacaa 1200
 gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagcccacgg tttgacgcct 1260
 gcacaagtgg tcgccatcgc ctccaatatt ggcggttaagc aggcgctgga aacagtacag 1320
 cgcctgctgc ctgtactgtg ccaggatcat ggactgacc cagaccagggt agtcgcaatc 1380
 gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt 1440
 tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa tgggggtggc 1500
 aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
 actcccgatc aagttgtagc gattgcgtcc aacgggtggag ggaaacaagc attggagact 1620
 gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
 gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740

```

gtactgtgcc aggatcatgg actgaccca gaccaggtag tgcgaatcgc gtcaaacgga      1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac      1860
ggccttacac cggagcaagt cgtggccatt gcatcccacg acggtggcaa acaggctctt      1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa      1980
gttgtagcga ttgcgtcgca tgacggaggg aaacaagcat tggagactgt ccaacggctc      2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccagc      2100
catgatggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag      2160
gatcatggac tgacacccga acaggtggtc gccattgctt cccacgacgg aggacggcca      2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg      2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa      2340
aagggtctgc ctcatgtctc cgcattgatc aaaagaacca accggcggat tcccgagaga      2400
acttcccatc gagtgcgagg atcc                                     2424

```

<210> 256

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 256

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 257

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 257

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac caccgagcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtagtcgc aatcgctca catgacgggg gaaagcaagc cctggaaacc	600
gtgcaaaggg tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660

```

gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat 780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac 840
ggtttgacgc ctgcacaagt ggtcgccatc gccaacaaca acggcggtaa gcaggcgctg 900
gaaacagtac agcgcctgct gcctgtactg tgccaggatc atggactgac cccagaccag 960
gtagtgcgaa tcgcgtcaca tgacggggga aagcaagccc tggaaaccgt gcaaaggttg 1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcatcc 1080
cacgacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
gccacggggc tgactcccg a tcaagttgta gcgattgcga ataacaatgg agggaaacaa 1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg tttgacgcct 1260
gcacaagtgg tcgccatcgc caacaacaac ggcggtgaagc aggcgctgga aacagtacag 1320
cgcctgtgtc ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc 1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt 1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaaataa taacgggtggc 1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaacaagc attggagact 1620
gtccaacggc tccttcccg gtgtgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcgaacatt 1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaagcaatg ggggtggcaa acaggctctt 1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gtttagcgga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggctctg acaccgaac aggtggctgc cattgcttct 2100
aacatcggag gacggccagc cttggagtcc atcgtagccc aattgtccag gcccgatccc 2160
gcgttggtg cgtaaacgaa tgaccatctg gtggcggttg catgtcttgg tggacgacct 2220
gcgctcgatg cagtcaaaaa gggctctgcct catgctcccg cattgatcaa aagaaccaac 2280
cggcggattc ccgagagaac ttcccatcga gtcgcgggat cc 2322

```

<210> 258

<211> 774

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 258

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His

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

275

280

285

Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495

Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 Page 111

[illegible]

```
<210> 259
<211> 2220
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polynucleotide
```

<400> 259

gctagcacca	tggactacaa	agaccatgac	ggtgattata	aagatcatga	catcgattac	60
aaggatgacg	atgacaagat	ggccccaag	aagaagagga	aggtgggcat	tcaccgcg	120
gtacctatgg	tggacttgag	gacactcggt	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaagggtca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggcttcact	240
catgcgcata	ttgtcgcgct	ttcacagcac	cctgcgggcg	ttgggacggt	ggctgtcaaa	300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcgg	360
aaacagtgg	cgggagcg	agcacttgag	gcgctgctga	ctgtggcg	tgagcttagg	420
gggcctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcgaagag	agggggagta	480
acagcggtag	aggcagtga	cgctggcg	aatgcgctca	ccggggcccc	cttgaacctg	540
acccagacc	aggtagtcgc	aatcgcgctg	aacattgggg	gaaagcaagc	cctggaaacc	600
gtgcaaaggt	tggtgccggt	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcaa	ataataacgg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gtcgcatgac	780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagcccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gccacaaca	acggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgcgtcgaa	cattggggga	aagcaagccc	tggaaaccgt	gcaaaggttg	1020
ttgccggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaaat	1080
aataacggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gccacgggc	tgactccga	tcaagttgta	gcgattgcgt	cgcatgacgg	agggaaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgttggtgc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	caacaacaac	ggcggttaagc	aggcgctgga	aacagtacag	1320
cgcctgctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccaggt	agtcgcaatc	1380
gcgtcacatg	acgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttggt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaaataa	taacgggtggc	1500
aaacaggctc	ttgagacggt	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgtcg	aacattggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gccacgggtt	tgacgcctgc	acaagtggtc	1680
gccatcgcca	gccatgatgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcacatgac	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcatcccacg	acgggtggcaa	acaggctcct	1920
gagacggttc	agagacttct	cccagttctc	tgtcaagccc	acgggctgac	acccgaacag	1980
gtggtcgcca	ttgcttccca	cgacggagga	cggccagcct	tggagtccat	cgtagcccaa	2040


```

ttgtccaggc cccatcccg gttggctgcg ttaacgaatg accatctggt ggcgttggca      2100
tgtcttgggtg gacgacccgc gctcgatgca gtcaaaaagg gtctgcctca tgctcccgc      2160
ttgatcaaaa gaaccaaccg gcggattccc gagagaactt cccatcgagt cgcgggatcc      2220

```

<210> 260
 <211> 740
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 polypeptide

<400> 260
 Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
 1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
 20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
 35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Arg Pro
 660 665 670
 Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu
 675 680 685
 Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly
 690 695 700
 Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro Ala
 705 710 715 720
 Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser His Arg
 725 730 735
 Val Ala Gly Ser
 740

<210> 261
 <211> 2424
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 261
 gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60
 aaggatgacg atgacaagat ggccccaag aagaagagga aggtgggcat tcaccgcggg 120
 gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag 180
 cctaagggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact 240
 catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa 300
 taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg 360
 aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420
 gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480
 acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg 540
 accccagacc aggtagtcgc aatcgcgtcg aacattgggg gaaagcaagc cctggaaacc 600
 gtgcaaaggt tgttgccggg cttttgtcaa gaccacggcc ttacaccgga gcaagtcgtg 660
 gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
 gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat 780
 ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagccac 840
 ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg 900
 gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac cccagaccag 960
 gtagtcgcaa tcgcgaacaa taatggggga aagcaagccc tggaaaccgt gcaaagggtg 1020
 ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcatcc 1080
 cacgacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
 gccacggg gcactccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa 1200
 gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg tttgacgcct 1260
 gcacaagtgg tcgcatcgc ctccaatatt ggcggtgaagc aggcgctgga aacagtacag 1320
 cgcctgtgtc ctgtactgtg ccaggatcat ggactgacct cagaccaggg agtcgcaatc 1380
 gcgtcgaaca ttgggggaaa gcaagccctg gaaaccgtgc aaagggtgtt gccggtcctt 1440
 tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaaataa taacggtggc 1500
 aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
 actcccgatc aagttgtagc gattgcgaat aacaatggag ggaaacaagc attggagact 1620
 gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680

```

gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgacccca gaccaggtag tcgcaatcgc gtcacatgac 1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgtccaa cgggtggagg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccaac 2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacaccga acaggtggtc gccattgctt ctaatggggg aggacggcca 2220
gccttgagat ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggtctgc ctcatgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gagtcgcggg atcc 2424

```

<210> 262

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 262

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 263
<211> 2424
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 263
gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60
aaggatgacg atgacaagat ggccccaag aagaagagga aggtgggcat tcaccgcggg 120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag 180
cctaaggtca ggagcaccgt cgcgcaacac cagagggcgc ttgtggggca tggcttcact 240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa 300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg 360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480
acagcggtag aggcagtga cgcctggcgc aatgcgctca cgggggcccc cttgaacctg 540
acccagacc aggtagtcgc aatcgcaac aataatgggg gaaagcaagc cctggaaacc 600
gtgcaaaggg tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg 660

29539_0036W01_SL2953936W.txt

```

gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgcatgac 780
ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac 840
ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg 900
gaaacagtac agcgcttgc tgcctgtactg tgccaggatc atggactgac cccagaccag 960
gtagtgcgaa tcgcgtcaca tgacggggga aagcaagccc tggaaaccgt gcaaaggttg 1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaatt 1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
gccacgggc tgactccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa 1200
gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagcccacgg tttgacgcct 1260
gcacaagtgg tcgccatcg caacaacaac ggcggtgaagc aggcgctgga aacagtacag 1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc 1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt 1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcatcca cgacggtggc 1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
actcccgatc aagttgtagc gattgcgaat aacaatggag ggaaacaagc attggagact 1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
gccatcgcca acaacaacgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcgaacatt 1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaaataata acggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgctgcaa cgggtggagg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgctgcac aagtggtcgc catcgctcg 2100
aatggcggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacaccga acaggtggtc gccattgctt cccacgacgg aggacggcca 2220
gccttgagat ccatcgtagc ccaattgtcc aggccgatc ccgcttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgctcga tgcagtcaaa 2340
aagggtctgc ctatgctcc cgcattgatc aaaagaacca accggcggtat tcccgagaga 2400
acttcccatc gagtcgagg atcc 2424

```

<210> 264

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 264

```

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1      5      10     15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
      20     25     30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
      35     40     45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
      50     55     60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
      65     70     75     80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
      85     90     95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
      100    105    110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
      115    120    125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
      130    135    140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
      145    150    155    160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
      165    170    175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
      180    185    190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
      210    215    220

Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
      225    230    235    240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
      245    250    255

```

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

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

Thr Ser His Arg Val Ala Gly Ser
805

<210> 265
<211> 2424
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

```

<400> 265
gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac      60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg      120
gtacctatgg tggacttgag gacactcggg tatttcgcaac agcaacagga gaaaatcaag      180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact      240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa      300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg      360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg      420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta      480
acagcggtag aggcatgca cgcctggcgc aatgcgctca ccggggcccc cttgaacctg      540
accccagacc aggtagtcgc aatcgcgaac aataatgggg gaaagcaagc cctggaaacc      600
gtgcaaaggt tggtgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg      660
gccattgcaa gcaacatcgg tggcaaacag gctcttgaga cggttcagag acttctccca      720
gttctctgtc aagccacggg gctgactccc gatcaagttg tagcgattgc gtcgaacatt      780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gccacaaca acggcggtaa gcaggcgctg      900
gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac ccagaccag      960
gtagtgcgaa tcgcgaacaa taatggggga aagcaagccc tggaaaccgt gcaaaggttg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat     1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gcccacgggc tgactcccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa     1200
gcattggaga ctgtccaacg gtccttccc gtgttgtgtc aagcccacgg tttgacgcct     1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtgaagc aggcgctgga aacagtacag     1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggg agtcgcaatc     1380
gcgtcgaaca ttgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa tgggggtggc     1500
aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaaacaagc attggagact     1620

```

```

gtccaacggc tccttcccggt gttgtgtcaa gcccacgggt tgacgcctgc acaagtggtc 1680
gccatcgcct ccaatattgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcacatgac 1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcatcccacg acggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgtccaa cgggtggagg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgctcg 2100
aatggcggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacaccga acagggtggtc gccattgctt ctaatggggg aggacggcca 2220
gccttgaggt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggctctg ctcattgtcc cgcattgatc aaaagaacca accggcggat tcccagaga 2400
acttcccatc gagtcgcggt atcc 2424

```

<210> 266

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 266

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
Page 127

115

120

125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220

Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255

Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 Page 128

385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 Page 129

660

665

670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 267

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 267

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60

aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg 120

gtacctatgg tggacttgag gacactcggt tattcgcaac agcaacagga gaaaatcaag 180

cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact 240

catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa 300

taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcgg 360

aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420

gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480

acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg 540

accccagacc aggtagtcgc aatcgcgtca catgacgggg gaaagcaagc cctggaaacc 600

```

gtgcaaaggt tgttgccggt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg      660
gccattgcaa gcaatggggg tggcaaacag gctcttgaga cggttcagag acttctccca      720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgaacatt      780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg      900
gaaacagtac agcgcctgct gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtcgcaa tcgctgcaaa cggaggggga aagcaagccc tggaaaccgt gcaaaggttg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat     1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gccacggggc tgactccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg ttgacgcct      1260
gcacaagtgg tcgccatcgc ctccaatatt ggcggtgaagc aggcgctgga aacagtacag     1320
cgcctgtctg ctgtactgtg ccaggatcat ggactgacc cagaccaggt agtcgcaatc     1380
gcgtcaaacg gagggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcatcca cgacggtggc     1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actccgatc aagttgtagc gattgcgtcc aacggtggag ggaacaagc attggagact     1620
gtccaacggc tccttcccg tttgtgtcaa gccacgggtt tgacgcctgc acaagtggctc     1680
gccatcgcct cgaatggcgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct     1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gaacaataat     1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac     1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt     1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa     1980
gttgtagcga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc     2040
cttcccgtgt tgtgtcaagc ccacggtctg acaccgaac aggtggctgc cattgcttcc     2100
cacgacggag gacggccagc cttggagtcc atcgtagccc aattgtccag gcccgatccc     2160
gcgttggtg cgtaaacgaa tgaccatctg gtggcggttg catgtcttgg tggacgaccc     2220
gcgctcgatg cagtcaaaaa gggctctgcct catgtctccg cattgatcaa aagaaccaac     2280
cggcggattc ccgagagaac ttcccatcga gtcgcgggat cc                        2322

```

<210> 268

<211> 774

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 268

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
210 215 220

Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
245 250 255

Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
260 265 270

Leu Pro Val₂₇₅ Leu Cys Gln Ala His₂₈₀ Gly Leu Thr Pro Ala₂₈₅ Gln Val Val
 Ala Ile₂₉₀ Ala Ser His Asp₂₉₅ Gly Gly Lys Gln Ala Leu₃₀₀ Glu Thr Val Gln
 Arg₃₀₅ Leu Leu Pro Val Leu₃₁₀ Cys Gln Asp His₃₁₅ Gly Leu Thr Pro Asp Gln₃₂₀
 Val Val Ala Ile₃₂₅ Ala Ser Asn Gly Gly₃₃₀ Lys Gln Ala Leu Glu Thr₃₃₅
 Val Gln Arg Leu₃₄₀ Leu Pro Val Leu₃₄₅ Cys Gln Asp His Gly Leu₃₅₀ Thr Pro
 Glu Gln Val₃₅₅ Val Ala Ile Ala₃₆₀ Asn Asn Asn Gly Gly Lys₃₆₅ Gln Ala Leu
 Glu Thr₃₇₀ Val Gln Arg Leu Leu₃₇₅ Pro Val Leu Cys Gln Ala His Gly Leu₃₈₀
 Thr₃₈₅ Pro Asp Gln Val Val₃₉₀ Ala Ile Ala Ser Asn₃₉₅ Ile Gly Gly Lys Gln₄₀₀
 Ala Leu Glu Thr Val₄₀₅ Gln Arg Leu Leu₄₁₀ Pro Val Leu Cys Gln Ala His₄₁₅
 Gly Leu Thr Pro₄₂₀ Ala Gln Val Val₄₂₅ Ala Ile Ala Ser Asn Ile Gly Gly₄₃₀
 Lys Gln Ala₄₃₅ Leu Glu Thr Val₄₄₀ Gln Arg Leu Leu Pro Val Leu Cys Gln₄₄₅
 Asp His₄₅₀ Gly Leu Thr Pro Asp₄₅₅ Gln Val Val Ala Ile₄₆₀ Ala Ser Asn Gly
 Gly Gly Lys Gln Ala Leu₄₇₀ Glu Thr Val Gln Arg Leu Leu Pro Val Leu₄₈₀
 Cys Gln Asp His Gly₄₈₅ Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser₄₉₅
 His Asp Gly Gly₅₀₀ Lys Gln Ala Leu Glu₅₀₅ Thr Val Gln Arg Leu Leu Pro₅₁₀
 Val Leu Cys₅₁₅ Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 Ala Ser Asn Gly Gly Gly Lys₅₃₅ Gln Ala Leu Glu Thr Val Gln Arg Leu₅₄₀
 Ala Ser₅₃₀

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
545 550 555 560

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
580 585 590

Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
610 615 620

Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
690 695 700

Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
705 710 715 720

Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu
725 730 735

Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala
740 745 750

Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser
755 760 765

His Arg Val Ala Gly Ser
770

<210> 269

<211> 2220

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 269

gctagcacca	tggaactaaa	agaccatgac	ggtgattata	aagatcatga	catcgattac	60
aaggatgacg	atgacaagat	ggccccaag	aagaagagga	aggtgggcat	tcaccgcggg	120
gtacctatgg	tggaacttag	gacactcgg	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaaggatca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggtttcact	240
catgcgcata	ttgtcgcgct	ttcacagcac	cctgcggcgc	ttgggacggt	ggctgtcaaa	300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcgg	360
aaacagtgg	cgaggcgcg	agcacttag	gcgctgctga	ctgtggcggg	tgagcttagg	420
gggcctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcgaagag	agggggagta	480
acagcggtag	aggcagtga	cgctggcgc	aatgcgctca	ccggggcccc	cttgaacctg	540
acccagacc	aggtagtcgc	aatcgcgctca	catgacgggg	gaaagcaagc	cctggaaacc	600
gtgcaaagg	tggtgccgg	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcaa	ataataacgg	tggaacaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gaataacaat	780
ggagggaaac	aagcattgga	gactgtccaa	cggtccttc	ccgtgtttgtg	tcaagcccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gccagccatg	atggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgcgtcaca	tgacggggga	aagcaagccc	tggaaccgt	gcaaaggttg	1020
ttgccgggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aacatcggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gccacgggc	tgactcccga	tcaagttgta	gcgattgcgt	cgcatgacgg	agggaaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgtttgtgtc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	cagccatgat	ggcggttaagc	aggcgctgga	aacagtacag	1320
cgctgtgtgc	ctgtactgtg	ccaggatcat	ggactgaccc	cagaccaggt	agtcgcaatc	1380
gcgtcgaaca	ttgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttgtt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaagcaa	tgggggtggc	1500
aaacaggctc	ttgagacgg	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgaat	aacaatggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gccacgggtt	tgacgcctgc	acaagtggtc	1680
gcatcgcct	cgaatggcgg	cggttaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcacatgac	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttgttgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcatcccacg	acggtggcaa	acaggctcct	1920

gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac acccgaacag 1980
 gtggtcgcca ttgcttccca cgacggagga cggccagcct tggagtccat cgtagcccaa 2040
 ttgtccaggc ccgatcccg cgttgctgcg ttaacgaatg accatctggt ggcgttggca 2100
 tgtcttgggtg gacgaccgc gctcgatgca gtcaaaaagg gtctgcctca tgctcccgca 2160
 ttgatcaaaa gaaccaaccg gcggattccc gagagaactt cccatcgagt cgcgggatcc 2220

<210> 270

<211> 740

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 270

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
 1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
 20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
 35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 450 455 460

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

Val Ala Gly Ser
740

<210> 271
<211> 2424
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 271
gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg 120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag 180
cctaaggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact 240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa 300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg 360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg 540
accccagacc aggtagtcgc aatcgcgtca catgacgggg gaaagcaagc cctggaaacc 600
gtgcaaaggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg 660
gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgaacatt 780
ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac 840
ggtttgacgc ctgcacaagt ggtcgccatc gccacaaca acggcggtaa gcaggcgctg 900
gaaacagtac agcgcctgct gcctgtactg tgccaggatc atggactgac cccagaccag 960
gtagtgcgaa tcgcgaacaa taatggggga aagcaagccc tggaaaccgt gcaaaggttg 1020
ttgccgggtc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcatcc 1080
cacgacggtg gaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
gccacggggc tgactcccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa 1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg tttgacgcct 1260
gcacaagtgg tcgccatcgc caacaacaac ggcggtaagc aggcgctgga aacagtacag 1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggg agtcgcaatc 1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaagggttgtt gccggtcctt 1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa tgggggtggc 1500
aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
actcccgatc aagttgtagc gattgcgaat aacaatggag ggaaacaagc attggagact 1620

```

gtccaacggc tccttcccggt gttgtgtcaa gcccacgggt tgacgcctgc acaagtggtc 1680
gccatcgcca acaacaacgg cggtgaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgacccca gaccaggtag tcgcaatcgc gtcgaacatt 1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaaataata acggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgtcgca tgacggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgccagc 2100
catgatggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacacccga acagggtggtc gccattgctt cccacgacgg aggacggcca 2220
gccttgagat ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggctctg ctcattgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gattcgcggt atcc 2424

```

<210> 272

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 272

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125
 Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 273

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 273

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcaagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540

acccagacc	aggtagtcgc	aatcgcgctca	catgacgggg	gaaagcaagc	cctggaaacc	600
gtgcaaaggt	tgttgccggt	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcaa	gcaatggggg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gaataacaat	780
ggagggaaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagcccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gcctccaata	ttggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgcgctcaa	cattggggga	aagcaagccc	tggaaaccgt	gcaaagggtt	1020
ttgccggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aacatcggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gcccacgggc	tgactccga	tcaagttgta	gcgattgcgt	cgaacattgg	agggaacaaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgtttgtgtc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	ctccaatatt	ggcggtaagc	aggcgctgga	aacagtacag	1320
cgctgtctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccaggt	agtcgcaatc	1380
gcgaacaata	atgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttggt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaagcaa	catcggtggc	1500
aaacaggctc	ttgagacggt	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgtcg	catgacggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gcccacgggt	tgacgcctgc	acaagtggtc	1680
gccatcgcct	cgaatggcgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcacatgac	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttgttgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcaagcaatg	ggggtggcaa	acaggctctt	1920
gagacggttc	agagacttct	cccagttctc	tgtcaagccc	acgggctgac	tcccgatcaa	1980
gttgtagcga	ttgcgaataa	caatggaggg	aaacaagcat	tggagactgt	ccaacggctc	2040
cttcccgtgt	tgtgtcaagc	ccacggtttg	acgcctgcac	aagtggtcgc	catcgccagc	2100
catgatggcg	gtaagcaggc	gctggaaaca	gtacagcgcc	tgctgcctgt	actgtgccag	2160
gatcatggac	tgacacccga	acagggtggc	gccattgctt	ctaacatcgg	aggacggcca	2220
gccttggagt	ccatcgtagc	ccaattgtcc	aggcccgatc	ccgcgttggc	tgcgttaacg	2280
aatgaccatc	tgggtggcgtt	ggcatgtctt	ggtggacgac	ccgcgctcga	tgacgtcaaa	2340
aagggtctgc	ctcatgtctc	cgcatgtatc	aaaagaacca	accggcggat	tcccgagaga	2400
acttcccatc	gagtcgctgg	atcc				2424

<210> 274

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 274

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
 1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
 20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
 35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Lys Ile Lys Pro Lys Val Arg
 50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220

Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 Page 145

245

250

255

Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 Page 146

515

520

525

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 Page 147

785

790

795

800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 275

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 275

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttact	240
catgcgcata ttgtcgcgt ttacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtagtcgc aatcggtca catgacgggg gaaagcaagc cctggaaacc	600
gtgcaaagggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgaacatt	780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgttggtg tcaagcccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gcctcgaatg gcggcggtaa gcaggcgctg	900
gaaacagtac agcgctgct gcctgtactg tgccaggatc atggactgac ccagaccag	960
gtagtcgcaa tcgctgcaaa cggaggggga aagcaagccc tggaaaccgt gcaaagggtg	1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc	1080
aatgggggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa	1140
gccacgggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgttggtgc aagcccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc ctcgaatggc ggcggttaagc aggcgctgga aacagtacag	1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacc cagaccaggg agtcgcaatc	1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaagggtgtt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa tgggggtggc	1500
aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560

```

actcccgatc aagttgtagc gattgcgtcg catgacggag ggaacaagc attggagact 1620
gtccaacggc tccttcccggt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
gccatcgcct ccaatattgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcaaacgga 1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcatccacg acggtggcaa acaggctctt 1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gtttagcga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgcctcg 2100
aatggcggcg gtaagcaggc gctggaaaca gtacagcgc tgctgcctgt actgtgccag 2160
gatcatggac tgacaccga acaggtggtc gccattgctt ctaacatcgg aggacggcca 2220
gccttgagc ccatcgtagc ccaattgtcc aggccgatc ccgcttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgctcga tgcagtcaaa 2340
aagggtctgc ctcatgtcc cgcattgatc aaaagaacca accggcggtat tcccgagaga 2400
acttcccatc gagtcgctgg atcc 2424

```

<210> 276

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 276

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125
 Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 277

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 277

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac caccgagcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480

acagcggtag	aggcagtgc	cgcctggcgc	aatgcgctca	ccggggcccc	cttgaacctg	540
accccagacc	aggtagtcgc	aatcgcgaa	aataatgggg	gaaagcaagc	cctggaaacc	600
gtgcaaaggt	tgttgccggt	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcaa	gcaacatcgg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gtcgaacatt	780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagcccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gcctccaata	ttggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgcgtcgaa	cattggggga	aagcaagccc	tggaaaccgt	gcaaagggtg	1020
ttgccggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aatgggggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gccacgggc	tgactccga	tcaagttgta	gcgattgcga	ataacaatgg	agggaaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgtttgtgtc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	ctccaatatt	ggcggttaagc	aggcgctgga	aacagtacag	1320
cgcctgctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccaggt	agtcgcaatc	1380
gcgtcacatg	acgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttggt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaagcaa	tgggggtggc	1500
aaacaggctc	ttgagacggt	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgaat	aacaatggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gccacgggtt	tgacgcctgc	acaagtggtc	1680
gccatcgcct	ccaatattgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcgaacatt	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcaagcaatg	ggggtggcaa	acaggctctt	1920
gagacgggtc	agagacttct	cccagttctc	tgtcaagccc	acgggctgac	tcccgatcaa	1980
gttgtagcga	ttgcgtcgaa	cattggaggg	aaacaagcat	tggagactgt	ccaacggctc	2040
cttcccgtgt	tgtgtcaagc	ccacggtttg	acgcctgcac	aagtggctgc	catcgcctcg	2100
aatggcggcg	gtaagcaggc	gctggaaaca	gtacagcgcc	tgctgcctgt	actgtgccag	2160
gatcatggac	tgacacccga	acaggtggtc	gccattgctt	ctaacatcgg	aggacggcca	2220
gccttggagt	ccatcgtagc	ccaattgtcc	aggcccgatc	ccgcgttggc	tgcgttaacg	2280
aatgaccatc	tgggtggcgtt	ggcatgtctt	ggtggacgac	ccgcgctcga	tgcagtcaaa	2340
aagggtctgc	ctcatgctcc	cgcattgatc	aaaagaacca	accggcggat	tcccgagaga	2400
acttcccatc	gagtcgcggg	atcc				2424

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 278

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

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

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

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 279

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 279

```

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac      60
aaggatgacg atgacaagat ggccccaag aagaagagga aggtgggcat tcaccgcggg      120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag      180
cctaagggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact      240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa      300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg      360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg      420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta      480
acagcggtag aggcagtgc cgcctggcgc aatgcgctca ccggggcccc cttgaacctg      540
acccagacc aggtagtcgc aatcgcgtca aacggagggg gaaagcaagc cctggaaacc      600
gtgcaaaggt tggtgccgt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg      660
gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca      720
gttctctgtc aagccacgg gctgactccc gatcaagttg tagcgattgc gtcgcatgac      780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgttggt tcaagccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg      900
gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtcgcaa tcgcgtcaaa cggaggggga aagcaagccc tggaaaccgt gcaaagggtg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc     1080
aacatcggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gcccacgggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgttggtgtc aagccacgg tttgacgcct     1260
gcacaagtgg tcgccatcgc caacaacaac ggcggtgaagc aggcgctgga aacagtacag     1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccagggt agtcgcaatc     1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcatccca cgacgggtggc     1500

```

```

aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg      1560
actccccgatc aagttgtagc gattgcgtcg aacattggag ggaacaagc attggagact      1620
gtccaacggc tccttcccggt gttgtgtcaa gccacagggt tgacgcctgc acaagtggctc      1680
gccatcgcca gccatgatgg cggttaagcag gcgctggaaa cagtacagcg cctgctgcct      1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcacatgac      1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac      1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt      1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa      1980
gtttagcgca ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc      2040
cttcccgtgt tgtgtcaagc ccacggctctg acaccgaac aggtggctcg cattgcttcc      2100
cacgacggag gacggccagc cttggagtcc atcgtagccc aattgtccag gcccgatccc      2160
gcgttggtctg cgtaaacgaa tgaccatctg gtggcggttg catgtcttgg tggacgaccc      2220
gcgctcgatg cagtcaaaaa gggctctgcct catgctcccg cattgatcaa aagaaccaac      2280
cggcggattc ccgagagaac ttcccatcga gtcgcgggat cc                        2322

```

<210> 280

<211> 774

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 280

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125
 Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140
 Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
675 680 685

Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
690 695 700

Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
705 710 715 720

Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu
725 730 735

Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala
740 745 750

Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser
755 760 765

His Arg Val Ala Gly Ser
770

<210> 281

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 281

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggt tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggt	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtgc cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtagtcgc aatcgcgaa aataatgggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcat ccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagccacgg gctgactccc gatcaagttg tagcgattgc gtccaacggt	780


```

ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgttggtg tcaagcccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gcctcgaatg gcggcggtaa gcaggcgctg      900
gaaacagtac agcgctgtct gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtgcgaa tcgctgcgaa cattggggga aagcaagccc tggaaaccgt gcaaagggttg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat     1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gccacgggc tgactcccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgttggtgc aagcccacgg tttgacgcct     1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtgaagc aggcgctgga aacagtacag     1320
cgctgtctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggc agtcgcaatc     1380
gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaagggttgtt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcatcca cgacggtggc     1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actcccgatc aagttgtagc gattgctgcc aacggtggag ggaaacaagc attggagact     1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc     1680
gccatcgcca acaacaacgg cggtaaagcag gcgctggaaa cagtacagcg cctgctgcct     1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gaacaataat     1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac     1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt     1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa     1980
gttgtagcga ttgctgcaa cggtggaggg aaacaagcat tggagactgt ccaacggctc     2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgctcg     2100
aatggcggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag     2160
gatcatggac tgacaccca acaggtggtc gccattgctt ctaatggggg aggacggcca     2220
gccttgagat ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg     2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa     2340
aagggtctgc ctcatgtctc cgcattgatc aaaagaacca accggcggat tcccgagaga     2400
acttcccatc gagtgcgagg atcc                                     2424

```

<210> 282

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 282

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His

```

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

```

275

280

285

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 Page 164

545 550 555 560
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 690 695 700
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720
 Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly
 725 730 735
 Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750
 Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765
 Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780
 His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800
 Thr Ser His Arg Val Ala Gly Ser
 805

<210> 283

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 283

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cagcaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
accccagacc aggtagtcgc aatcgcgtca aacggagggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat	780
ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gccaacaaca acggcggtaa gcaggcgctg	900
gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac cccagaccag	960
gtagtgcgaa tcgctgcaaa cggaggggga aagcaagccc tggaaaccgt gcaaaggttg	1020
ttgccgggtc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat	1080
aataacggtg gcaaacaggc tcttgagacg gttagagac tctctccagt tctctgtcaa	1140
gcccacgggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc ctcgaatggc ggcggtgaagc aggcgctgga aacagtacag	1320
cgcctgtgtc ctgtactgtg ccaggatcat ggactgaccc cagaccaggg agtcgcaatc	1380
gcgtcaaacg gagggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa catcggtggc	1500
aaacaggctc ttgagacggg tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaaacaagc attggagact	1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtgggc	1680
gccatcgcca gccatgatgg cggtaaagcag gcgctggaaa cagtacagcg cctgctgcct	1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcaaacgga	1800

```

gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaaataata acggtggcaa acaggctctt 1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gttgtagcga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgctcg 2100
aatggcggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacacccga acaggtggtc gccattgctt ctaatggggg aggacggcca 2220
gccttgagat ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggctctg ctcattgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gattcgcggg atcc 2424

```

<210> 284

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 284

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160
 Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 285

<211> 2322

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 285

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
accccagacc aggtatgcgc aatcgcgtca catgacgggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggg cttttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca	720

```

gttctctgtc aagccacagg gctgactccc gatcaagttg tagcgattgc gtcgcatgac      780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gcctccaata ttggcggtaa gcaggcgctg      900
gaaacagtac agcgcttgcg gcctgtactg tgccaggatc atggactgac cccagaccag      960
gtagtgcgaa tcgcgaacaa taatggggga aagcaagccc tggaaaccgt gcaaaggttg     1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc     1080
aacatcggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa     1140
gccacggggc tgactcccga tcaagttgta gcgattgcgt cgcattgacg agggaaacaa     1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagccacagg tttgacgcct     1260
gcacaagtgg tcgccatcgc ctccaatatt ggcggttaagc aggcgctgga aacagtacag     1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc     1380
gcgtcaaacg gagggggaaa gcaagccctg gaaaccgtgc aaaggttggt gccggtcctt     1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaaataa taacggtggc     1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg     1560
actcccgatc aagttgtagc gattgcgtcg aacattggag ggaacaagc attggagact     1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc     1680
gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct     1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcgaacatt     1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac     1860
ggccttacac cggagcaagt cgtggccatt gcaaataata acggtggcaa acaggctctt     1920
gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa     1980
gttgtagcga ttgcgtcgca tgacggaggg aaacaagcat tggagactgt ccaacggctc     2040
cttcccgtgt tgtgtcaagc ccacggtctg acaccgaac aggtggctgc cattgcttcc     2100
cacgacggag gacggccagc cttggagtcc atcgtagccc aattgtccag gcccgatccc     2160
gcgttggtcg cgtaacgaa tgaccatctg gtggcgttgg catgtcttgg tggacgacct     2220
gcgctcgatg cagtcaaaaa gggctctgcct catgtctccg cattgatcaa aagaaccaac     2280
cggcggattc ccgagagaac ttcccatcga gtcgcgggat cc                        2322

```

<210> 286

<211> 774

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 286

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

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

Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 690 695 700

Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
 705 710 715 720

Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu
 725 730 735

Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala
 740 745 750

Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg Thr Ser
 755 760 765

His Arg Val Ala Gly Ser
 770

<210> 287

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 287

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac

60

29539_0036W01_SL2953936W.txt

aaggatgacg	atgacaagat	ggcccccaag	aagaagagga	aggtgggcat	tcaccgcggg	120
gtacctatgg	tggacttgag	gacactcggg	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaagggtca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggtttcact	240
catgcgcata	ttgtcgcgt	ttcacagcac	cctgcggcgc	ttgggacggg	ggctgtcaaa	300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcggg	360
aaacagtggg	cgggagcgcg	agcacttgag	gcgctgctga	ctgtggcggg	tgagcttagg	420
gggcctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcgaagag	agggggagta	480
acagcggtag	aggcagtga	cgctggcg	aatgcgctca	ccggggcccc	cttgaacctg	540
acccagacc	aggtagtcgc	aatcgcgctca	catgacgggg	gaaagcaagc	cctggaaacc	600
gtgcaaagg	tgttgccgg	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcat	cccacgacgg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gtccaacggg	780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagcccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gcctcgaatg	gcggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgctgcaaa	cggaggggga	aagcaagccc	tggaaaccgt	gcaaagggtg	1020
ttgccgggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aatgggggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gcccacgggc	tgactcccga	tcaagttgta	gcgattgcga	ataacaatgg	agggaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgtttgtgtc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	ctcgaatggc	ggcggtaagc	aggcgctgga	aacagtacag	1320
cgctgtctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccagg	agtcgcaatc	1380
gcgtcaaacg	gagggggaaa	gcaagccctg	gaaaccgtgc	aaaggttgtt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaagcaa	tgggggtggc	1500
aaacaggctc	ttgagacggg	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgtcg	catgacggag	ggaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gcccacgggt	tgacgcctgc	acaagtgggc	1680
gccatcgcct	cgaatggcgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gaacaataat	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcacccacg	acgggtggcaa	acaggctctt	1920
gagacgggtc	agagacttct	cccagttctc	tgtcaagccc	acgggctgac	tcccgatcaa	1980
gttgtagcga	ttgcgtccaa	cgggtggagg	aaacaagcat	tggagactgt	ccaacggctc	2040
cttcccgtgt	tgtgtcaagc	ccacggtttg	acgcctgcac	aagtggtcgc	catcgcctcc	2100

```

aatattggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag      2160
gatcatggac tgacacccga acaggtggtc gccattgctt ctaacatcgg aggacggcca      2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg      2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa      2340
aagggctctg ctcattgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga      2400
acttcccatc gagtcgcggg atcc                                             2424

```

<210> 288

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 288

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255

Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590

Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Ile
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 289
 <211> 2424
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 289
 gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac 60
 aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg 120
 gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag 180
 cctaagggtca ggagcaccgt cgcgcaacac caccgagcgc ttgtggggca tggcttcact 240
 catgcgcata ttgtcgcgt ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa 300
 taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggg 360
 aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg 420
 gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta 480
 acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg 540
 accccagacc aggtagtcgc aatcgcgtca aacggagggg gaaagcaagc cctggaaacc 600
 gtgcaaagggt tggtgccggg ctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg 660
 gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca 720
 gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgcatgac 780
 ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgttgtg tcaagcccac 840
 ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg 900
 gaaacagtac agcgctgct gcctgtactg tgccaggatc atggactgac cccagaccag 960
 gtagtcgcaa tcgcgaacaa taatggggga aagcaagccc tggaaaccgt gcaaagggtg 1020

```

ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc 1080
aatgggggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
gccacggggc tgactcccgga tcaagttgta gcgattgcgt cgcattgacgg agggaaacaa 1200
gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagcccacgg tttgacgcct 1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtaagc aggcgctgga aacagtacag 1320
cgcctgtctg ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc 1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt 1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa catcgggtggc 1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
actcccgatc aagttgtagc gattgcgtcg aacattggag ggaacaagc attggagact 1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
gccatcgcca acaacaacgg cggttaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcacatgac 1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt 1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
gtttagcgga ttgcgtcgaa cattggaggg aaacaagcat tggagactgt ccaacggctc 2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgctcg 2100
aatggcgggc gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
gatcatggac tgacacccga acaggtggtc gccattgcta ataataacgg aggacggcca 2220
gccttggagt ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg 2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
aagggtctgc ctcatgtcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
acttcccatc gagtcgcggg atcc 2424

```

<210> 290

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 290

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr

35

40

45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 210 215 220

Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 Page 181

305				310				315				320			
Val	Val	Ala	Ile	Ala 325	Asn	Asn	Asn	Gly	Gly 330	Lys	Gln	Ala	Leu	Glu 335	Thr
Val	Gln	Arg	Leu 340	Leu	Pro	Val	Leu	Cys 345	Gln	Asp	His	Gly	Leu 350	Thr	Pro
Glu	Gln	Val 355	Val	Ala	Ile	Ala	Ser 360	Asn	Gly	Gly	Gly	Lys 365	Gln	Ala	Leu
Glu	Thr 370	Val	Gln	Arg	Leu	Leu 375	Pro	Val	Leu	Cys	Gln 380	Ala	His	Gly	Leu
Thr 385	Pro	Asp	Gln	Val	Val 390	Ala	Ile	Ala	Ser	His 395	Asp	Gly	Gly	Lys	Gln 400
Ala	Leu	Glu	Thr	Val 405	Gln	Arg	Leu	Leu	Pro 410	Val	Leu	Cys	Gln	Ala 415	His
Gly	Leu	Thr	Pro 420	Ala	Gln	Val	Val	Ala 425	Ile	Ala	Ser	His	Asp 430	Gly	Gly
Lys	Gln	Ala 435	Leu	Glu	Thr	Val	Gln 440	Arg	Leu	Leu	Pro	Val 445	Leu	Cys	Gln
Asp	His 450	Gly	Leu	Thr	Pro	Asp 455	Gln	Val	Val	Ala	Ile 460	Ala	Ser	His	Asp
Gly 465	Gly	Lys	Gln	Ala	Leu 470	Glu	Thr	Val	Gln	Arg 475	Leu	Leu	Pro	Val	Leu 480
Cys	Gln	Asp	His	Gly 485	Leu	Thr	Pro	Glu	Gln 490	Val	Val	Ala	Ile	Ala 495	Ser
Asn	Ile	Gly	Gly 500	Lys	Gln	Ala	Leu	Glu 505	Thr	Val	Gln	Arg	Leu 510	Leu	Pro
Val	Leu	Cys 515	Gln	Ala	His	Gly	Leu 520	Thr	Pro	Asp	Gln	Val 525	Val	Ala	Ile
Ala	Ser 530	Asn	Ile	Gly	Gly	Lys 535	Gln	Ala	Leu	Glu	Thr 540	Val	Gln	Arg	Leu
Leu 545	Pro	Val	Leu	Cys	Gln 550	Ala	His	Gly	Leu	Thr 555	Pro	Ala	Gln	Val	Val 560
Ala	Ile	Ala	Asn	Asn 565	Asn	Gly	Gly	Lys	Gln 570	Ala	Leu	Glu	Thr	Val 575	Gln
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln

580

585

590

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 291

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 291

gctagcacca	tggactacaa	agaccatgac	ggtgattata	aagatcatga	catcgattac	60
aaggatgacg	atgacaagat	ggcccccaag	aagaagagga	aggtgggcat	tcaccgcggg	120
gtacctatgg	tggacttgag	gacactcggt	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaagggtca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggcttcact	240
catgcgcata	ttgtcgcgct	ttcacagcac	cctgcggcgc	ttgggacggt	ggctgtcaaa	300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcgggt	360
aaacagtgggt	cgggagcgcg	agcacttgag	gcgctgctga	ctgtggcggg	tgagcttagg	420
gggcctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcgaagag	agggggagta	480
acagcggtag	aggcagtga	cgcctggcgc	aatgcgctca	ccggggcccc	cttgaacctg	540
acccagacc	aggtagtcgc	aatcgcgaa	aataatggg	gaaagcaagc	cctggaaacc	600
gtgcaaaggt	tggtgccggt	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg	660
gccattgcaa	gcaatggggg	tggcaaacag	gctcttgaga	cggttcagag	acttctccca	720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gtccaacggt	780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagccac	840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gccagccatg	atggcggtaa	gcaggcgctg	900
gaaacagtac	agcgctgct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag	960
gtagtcgcaa	tcgcgtcgaa	cattggggga	aagcaagccc	tggaaaccgt	gcaaagggtg	1020
ttgccggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcaagc	1080
aacatcggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa	1140
gcccacgggc	tgactccga	tcaagttgta	gcgattgcgt	ccaacgggtg	agggaaacaa	1200
gcattggaga	ctgtccaacg	gctccttccc	gtgttggtgc	aagcccacgg	tttgacgcct	1260
gcacaagtgg	tcgccatcgc	ctccaatatt	ggcggttaagc	aggcgctgga	aacagtacag	1320
cgcctgctgc	ctgtactgtg	ccaggatcat	ggactgacct	cagaccaggt	agtcgcaatc	1380
gcgtcaaacg	gagggggaaa	gcaagccctg	gaaaccgtgc	aaaggttgtt	gccggtcctt	1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcatccca	cgacgggtggc	1500
aaacaggctc	ttgagacggt	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg	1560
actcccgatc	aagttgtagc	gattgcgaat	aacaatggag	ggaaacaagc	attggagact	1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gcccacgggt	tgacgcctgc	acaagtggctc	1680
gccatcgcct	cgaatggcgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct	1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcacatgac	1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac	1860
ggccttacac	cggagcaagt	cgtggccatt	gcatccacag	acgggtggcaa	acaggctcct	1920
gagacgggtc	agagacttct	cccagttctc	tgtcaagccc	acgggctgac	tcccgatcaa	1980
gttgtagcga	ttgcgaataa	caatggaggg	aaacaagcat	tggagactgt	ccaacggctc	2040

```

cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccaac      2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag      2160
gatcatggac tgacacccga acaggtggtc gccattgcta ataataacgg aggacggcca      2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg      2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa      2340
aagggctctgc ctcatgctcc cgcattgatc aaaagaacca accggcggat tcccgagaga      2400
acttcccatc gagtcgctgg atcc                                          2424

```

<210> 292

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 292

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 690 695 700
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 293

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 293

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggt tattcgcaac agcaacagga gaaaatcaag	180
cctaagggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga ggcacgcacg aggcaattgt aggggtcgggt	360
aaacagtgggt cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcagtga cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
acccagacc aggtagtcgc aatcgcgaa aataatgggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcaa ataataacgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gaataacaat	780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg	900
gaaacagtac agcgctgct gcctgtactg tgccaggatc atggactgac cccagaccag	960

gtagtcgcaa tcgcgtcaca tgacggggga aagcaagccc tggaaaccgt gcaaagggtt 1020
 ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaagc 1080
 aacatcggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa 1140
 gccacgggc tgactcccga tcaagttgta gcgattgcga ataacaatgg agggaaacaa 1200
 gcattggaga ctgtccaacg gctccttccc gtgttggtgc aagcccacgg tttgacgcct 1260
 gcacaagtgg tcgccatcgc ctccaatatt ggcggtaagc aggcgctgga aacagtacag 1320
 cgctgtctgc ctgtactgtg ccaggatcat ggactgacct cagaccaggt agtcgcaatc 1380
 gcgaacaata atgggggaaa gcaagccctg gaaaccgtgc aaagggttgtt gccggtcctt 1440
 tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa catcggtggc 1500
 aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg 1560
 actcccgatc aagttgtagc gattgcgtcc aacggtggag ggaaacaagc attggagact 1620
 gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc 1680
 gccatcgcca acaacaacgg cggttaagcag gcgctggaaa cagtacagcg cctgctgcct 1740
 gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gaacaataat 1800
 gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac 1860
 ggccttacac cggagcaagt cgtggccatt gcatccacg acggtggcaa acaggctctt 1920
 gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa 1980
 gttgtagcga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc 2040
 cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgccaac 2100
 aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag 2160
 gatcatggac tgacaccca acaggtggtc gccattgctt cccacgacgg aggacggcca 2220
 gccttgagat ccatcgtagc ccaattgtcc aggccgatc ccgcgttggc tgcgttaacg 2280
 aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa 2340
 aagggtctgc ctcatgctcc cgcattgatc aaaagaacca accggcggat tcccgagaga 2400
 acttcccatc gagtcgcggg atcc 2424

<210> 294

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 294

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
 1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
 20 25 30

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

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 420 425 430
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590

Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 295

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 295

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac	60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg	120
gtacctatgg tggacttgag gacactcggt tattcgcaac agcaacagga gaaaatcaag	180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact	240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggt ggctgtcaaa	300
taccaagata tgattgcggc cctgcccga gccacgcacg aggcaattgt aggggtcggt	360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg	420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta	480
acagcggtag aggcatgca cgcctggcgc aatgcgctca ccggggcccc cttgaacctg	540
accccagacc aggtatgcgc aatcgcgtcg aacattgggg gaaagcaagc cctggaaacc	600
gtgcaaaggt tgttgccggt cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg	660
gccattgcaa gcaacatcgg tggcaaacag gctcttgaga cggttcagag acttctccca	720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgaacatt	780
ggagggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac	840
ggtttgacgc ctgcacaagt ggtcgccatc gccaacaaca acggcggtaa gcaggcgctg	900
gaaacagtac agcgctgtgt gcctgtactg tgccaggatc atggactgac cccagaccag	960
gtagtcgcaa tcgcgtcaca tgacggggga aagcaagccc tggaaaccgt gcaaaggttg	1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcatcc	1080
cacgacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa	1140
gcccacgggc tgactcccga tcaagttgta gcgattgcga ataacaatgg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgttgtgtc aagcccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtaagc aggcgctgga aacagtacag	1320
cgcctgtgtc ctgtactgtg ccaggatcat ggactgacct cagaccagggt agtcgcaatc	1380
gcgtcacatg acgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaaataa taacggtggc	1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560
actcccgatc aagttgtagc gattgcgtcg catgacggag ggaaacaagc attggagact	1620
gtccaacggc tccttcccgt gttgtgtcaa gccacgggtt tgacgcctgc acaagtggtc	1680
gccatcgcca gccatgatgg cggtaaagcag gcgctggaaa cagtacagcg cctgctgcct	1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcaaacgga	1800
gggggaaagc aagccctgga aaccgtgcaa aggttggtgc cggtcctttg tcaagaccac	1860
ggccttacac cggagcaagt cgtggccatt gcacccacg acggtggcaa acaggctctt	1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa	1980


```

gttgtagcga ttgcgtcgca tgacggaggg aaacaagcat tggagactgt ccaacggctc      2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgccaac      2100
aacaacggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag      2160
gatcatggac tgacacccga acaggtggtc gccattgcta ataataacgg aggacggcca      2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg      2280
aatgaccatc tgggtggcggtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa      2340
aagggctctgc ctcatgctcc cgcattgatc aaaagaacca accggcggat tcccagagaga      2400
acttcccatc gagtcgcggg atcc                                             2424

```

<210> 296

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 296

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn
 485 490 495
 Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Asn Asn Asn
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 297

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 297

```

gctagcacca tggactacaa agaccatgac ggtgattata aagatcatga catcgattac      60
aaggatgacg atgacaagat ggcccccaag aagaagagga aggtgggcat tcaccgcggg      120
gtacctatgg tggacttgag gacactcggg tattcgcaac agcaacagga gaaaatcaag      180
cctaaggtca ggagcaccgt cgcgcaacac cacgaggcgc ttgtggggca tggcttcact      240
catgcgcata ttgtcgcgct ttcacagcac cctgcggcgc ttgggacggg ggctgtcaaa      300
taccaagata tgattgcggc cctgcccga gacacgcacg aggcaattgt aggggtcggg      360
aaacagtggg cgggagcgcg agcacttgag gcgctgctga ctgtggcggg tgagcttagg      420
gggcctccgc tccagctcga caccgggcag ctgctgaaga tcgcgaagag agggggagta      480
acagcggtag aggcatgca cgcctggcgc aatgcgctca ccggggcccc cttgaacctg      540
accccagacc aggtagtcgc aatcgcaaac aataatgggg gaaagcaagc cctggaaacc      600
gtgcaaaggt tggtgccggg cctttgtcaa gaccacggcc ttacaccgga gcaagtcgtg      660
gccattgcat cccacgacgg tggcaaacag gctcttgaga cggttcagag acttctccca      720
gttctctgtc aagcccacgg gctgactccc gatcaagttg tagcgattgc gtcgcatgac      780
ggaggggaaac aagcattgga gactgtccaa cggctccttc ccgtgtttgtg tcaagcccac      840
ggtttgacgc ctgcacaagt ggtcgccatc gccagccatg atggcggtaa gcaggcgctg      900

```

gaaacagtac agcgcttgct gcctgtactg tgccaggatc atggactgac cccagaccag	960
gtagtcgcaa tcgcgtcgaa cattggggga aagcaagccc tggaaaccgt gcaaagggtg	1020
ttgccggtcc tttgtcaaga ccacggcctt acaccggagc aagtcgtggc cattgcaaat	1080
aataacggtg gcaaacaggc tcttgagacg gttcagagac ttctcccagt tctctgtcaa	1140
gcccacgggc tgactcccga tcaagttgta gcgattgcgt cgaacattgg agggaaacaa	1200
gcattggaga ctgtccaacg gctccttccc gtgtttgtgtc aagcccacgg tttgacgcct	1260
gcacaagtgg tcgccatcgc cagccatgat ggcggtgaagc aggcgctgga aacagtacag	1320
cgcctgctgc ctgtactgtg ccaggatcat ggactgaccc cagaccagggt agtcgcaatc	1380
gcgtcgaaca ttgggggaaa gcaagccctg gaaaccgtgc aaaggttgtt gccggtcctt	1440
tgtcaagacc acggccttac accggagcaa gtcgtggcca ttgcaagcaa catcggtggc	1500
aaacaggctc ttgagacggt tcagagactt ctcccagttc tctgtcaagc ccacgggctg	1560
actcccgatc aagttgtagc gattgcgaat aacaatggag ggaaacaagc attggagact	1620
gtccaacggc tccttcccgt gttgtgtcaa gcccacgggt tgacgcctgc acaagtggtc	1680
gccatcgcca gccatgatgg cggtaagcag gcgctggaaa cagtacagcg cctgctgcct	1740
gtactgtgcc aggatcatgg actgaccca gaccaggtag tcgcaatcgc gtcgaacatt	1800
gggggaaagc aagccctgga aaccgtgcaa aggttgttgc cggtcctttg tcaagaccac	1860
ggccttacac cggagcaagt cgtggccatt gcaagcaaca tcggtggcaa acaggctctt	1920
gagacgggtc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa	1980
gttgtagcga ttgcgtcgca tgacggaggg aaacaagcat tggagactgt ccaacggctc	2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggctgc catcgctcc	2100
aatattggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag	2160
gatcatggac tgacacccga acagggtggtc gccattgctt ctaatggggg aggacggcca	2220
gccttgaggt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg	2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa	2340
aagggtctgc ctcatgctcc cgcattgatc aaaagaacca accggcggat tcccgagaga	2400
acttcccatc gagtgcggg atcc	2424

<210> 298

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 298

Ala	Ser	Thr	Met	Asp	Tyr	Lys	Asp	His	Asp	Gly	Asp	Tyr	Lys	Asp	His
1				5				10						15	

Asp	Ile	Asp	Tyr	Lys	Asp	Asp	Asp	Lys	Met	Ala	Pro	Lys	Lys	Lys	

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
 35 40 45

Leu Gly Tyr Ser Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
 100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
 130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
 145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 180 185 190

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

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 Page 199

290

295

300

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350

Glu Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
 355 360 365

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
 385 390 395 400

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 450 455 460

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480

Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495

Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510

Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525

Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540

Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 Page 200

565

570

575

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590

Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620

Glu Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
 625 630 635 640

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655

Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln
 660 665 670

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685

Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser Asn Gly
 725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
 740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
 755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
 770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
 785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
 805

<210> 299

<211> 2424

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 299

gctagcacca	tgga	tacaa	agaccatgac	ggtgattata	aagatcatga	catcgattac	60
aaggatgacg	atgacaagat	ggccccaag	aagaagagga	aggtgggcat	tcaccgcggg		120
gtacctatgg	tgga	cttgag	gacactcggg	tattcgcaac	agcaacagga	gaaaatcaag	180
cctaagggtca	ggagcaccgt	cgcgcaacac	cacgaggcgc	ttgtggggca	tggttctact		240
catgcgcata	ttgtcgcgt	ttcacagcac	cctgcggcgc	ttgggacggg	ggctgtcaaa		300
taccaagata	tgattgcggc	cctgcccga	gccacgcacg	aggcaattgt	aggggtcggg		360
aaacagtggg	cgggagcgcg	agcacttgag	gcgctgctga	ctgtggcggg	tgagcttagg		420
gggcctccgc	tccagctcga	caccgggcag	ctgctgaaga	tcgcgaagag	agggggagta		480
acagcggtag	aggcagtga	cgctggcgc	aatgcgctca	ccggggcccc	cttgaacctg		540
acccagacc	aggtagtcgc	aatcgcgctca	catgacgggg	gaaagcaagc	cctggaaacc		600
gtgcaaagg	tgttgccggg	cctttgtcaa	gaccacggcc	ttacaccgga	gcaagtcgtg		660
gccattgcat	cccacgacgg	tggaacaacag	gctcttgaga	cggttcagag	acttctccca		720
gttctctgtc	aagcccacgg	gctgactccc	gatcaagttg	tagcgattgc	gtcgcatgac		780
ggagggaaac	aagcattgga	gactgtccaa	cggctccttc	ccgtgtttgtg	tcaagcccac		840
ggtttgacgc	ctgcacaagt	ggtcgccatc	gccagccatg	atggcggtaa	gcaggcgctg		900
gaaacagtac	agcgctgtct	gcctgtactg	tgccaggatc	atggactgac	cccagaccag		960
gtagtcgcaa	tcgcgaacaa	taatggggga	aagcaagccc	tggaaccgt	gcaaagggtg		1020
ttgccgggtcc	tttgtcaaga	ccacggcctt	acaccggagc	aagtcgtggc	cattgcatcc		1080
cacgacggtg	gcaaacaggc	tcttgagacg	gttcagagac	ttctcccagt	tctctgtcaa		1140
gccacgggc	tgactcccga	tcaagttgta	gcgattgcga	ataacaatgg	agggaacaa		1200
gcattggaga	ctgtccaacg	gctccttccc	gtgtttgtgtc	aagcccacgg	tttgacgcct		1260
gcacaagtgg	tcgccatcgc	caacaacaac	ggcggttaagc	aggcgctgga	aacagtacag		1320
cgctgtctgc	ctgtactgtg	ccaggatcat	ggactgaccc	cagaccagg	agtcgcaatc		1380
gcgtcacatg	acgggggaaa	gcaagccctg	gaaaccgtgc	aaaggttggt	gccggtcctt		1440
tgtcaagacc	acggccttac	accggagcaa	gtcgtggcca	ttgcaagcaa	tgggggtggc		1500
aaacagggtc	ttgagacggg	tcagagactt	ctcccagttc	tctgtcaagc	ccacgggctg		1560
actcccgatc	aagttgtagc	gattgcgtcg	catgacggag	ggaacaagc	attggagact		1620
gtccaacggc	tccttcccgt	gttgtgtcaa	gccacgggtt	tgacgcctgc	acaagtggtc		1680
gccatcgcca	gccatgatgg	cggtaagcag	gcgctggaaa	cagtacagcg	cctgctgcct		1740
gtactgtgcc	aggatcatgg	actgaccca	gaccaggtag	tcgcaatcgc	gtcacatgac		1800
gggggaaagc	aagccctgga	aaccgtgcaa	aggttggtgc	cggtcctttg	tcaagaccac		1860
ggccttacac	cggagcaagt	cgtggccatt	gcatcccacg	acggtggcaa	acaggctctt		1920

```

gagacggttc agagacttct cccagttctc tgtcaagccc acgggctgac tcccgatcaa      1980
gtttagcgga ttgcgaataa caatggaggg aaacaagcat tggagactgt ccaacggctc      2040
cttcccgtgt tgtgtcaagc ccacggtttg acgcctgcac aagtggtcgc catcgccagc      2100
catgatggcg gtaagcaggc gctggaaaca gtacagcgcc tgctgcctgt actgtgccag      2160
gatcatggac tgacacccga acaggtggtc gccattgctt cccacgacgg aggacggcca      2220
gccttggagt ccatcgtagc ccaattgtcc aggcccgatc ccgcgttggc tgcgttaacg      2280
aatgaccatc tgggtggcgtt ggcatgtctt ggtggacgac ccgcgctcga tgcagtcaaa      2340
aagggtctgc ctcatgtcc cgcatcgatc aaaagaacca accggcggat tcccgagaga      2400
acttcccatc gagtcgcggg atcc                                           2424

```

<210> 300

<211> 808

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 300

Ala Ser Thr Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His
1 5 10 15

Asp Ile Asp Tyr Lys Asp Asp Asp Asp Lys Met Ala Pro Lys Lys Lys
20 25 30

Arg Lys Val Gly Ile His Arg Gly Val Pro Met Val Asp Leu Arg Thr
35 40 45

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
50 55 60

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
65 70 75 80

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
85 90 95

Val Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr
100 105 110

His Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
115 120 125

Leu Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu
130 135 140

Gln Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val
145 150 155 160

Thr Ala Val Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala
 165 170 175
 Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 180 185 190
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 195 200 205
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 210 215 220
 His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 225 230 235 240
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 245 250 255
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 260 265 270
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 275 280 285
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 290 295 300
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 305 310 315 320
 Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr
 325 330 335
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 340 345 350
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 355 360 365
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 370 375 380
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 385 390 395 400
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 405 410 415
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 420 425 430

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 435 440 445
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
 450 455 460
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 465 470 475 480
 Cys Gln Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser
 485 490 495
 Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 500 505 510
 Val Leu Cys Gln Ala His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 515 520 525
 Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 530 535 540
 Leu Pro Val Leu Cys Gln Ala His Gly Leu Thr Pro Ala Gln Val Val
 545 550 555 560
 Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 565 570 575
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 580 585 590
 Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 595 600 605
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 610 615 620
 Glu Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 625 630 635 640
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His Gly Leu
 645 650 655
 Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 660 665 670
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala His
 675 680 685
 Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly
 690 695 700

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
705 710 715 720

Asp His Gly Leu Thr Pro Glu Gln Val Val Ala Ile Ala Ser His Asp
725 730 735

Gly Gly Arg Pro Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro
740 745 750

Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala
755 760 765

Cys Leu Gly Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro
770 775 780

His Ala Pro Ala Leu Ile Lys Arg Thr Asn Arg Arg Ile Pro Glu Arg
785 790 795 800

Thr Ser His Arg Val Ala Gly Ser
805

<210> 301
<211> 88
<212> DNA
<213> Homo sapiens

<400> 301
ttcccagact cagtgggaag agctccctca ccatgagtag cgctatgttg gtgacttgcc 60
tcccggaccc cagcagcagc ttccgtga 88

<210> 302
<211> 80
<212> DNA
<213> Homo sapiens

<400> 302
ttcccagact cagtgggaag agctccctca ccatgagtag cgctatgttg cctcccgga 60
cccagcagca gcttccgtga 80

<210> 303
<211> 57
<212> DNA
<213> Homo sapiens

<400> 303
ttcccagact cagtgggaag agctccctca cccggacccc agcagcagct tccgtga 57

<210> 304
<211> 88
<212> DNA
<213> Homo sapiens

<400> 304
ggcgtgggag agtggatttc cgaagctgac agatgggtat tctttgacgg ggggtagggg 60

cggaacctga gaggcgtaag gcgttgtg 88

<210> 305
 <211> 71
 <212> DNA
 <213> Homo sapiens

<400> 305
 ggcgtgggag agtggatttc cgaagctgac agatgggtag gggcggaacc tgagaggcgt 60
 aaggcgttgt g 71

<210> 306
 <211> 65
 <212> DNA
 <213> Homo sapiens

<400> 306
 ggcgtgggag agtggatttc cgaagctgac agaggggtag aacctgagag gcgtaaggcg 60
 ttgtg 65

<210> 307
 <211> 61
 <212> DNA
 <213> Homo sapiens

<400> 307
 ggcgtgggag agtggatttc cgaagctgac agatggaacc tgagaggcgt aaggcgttgt 60
 g 61

<210> 308
 <211> 50
 <212> DNA
 <213> Homo sapiens

<400> 308
 ggcgtgggag agtggatttc cgaagctgac agatggggta aggcgttgtg 50

<210> 309
 <211> 49
 <212> DNA
 <213> Homo sapiens

<400> 309
 ggcgtgggag agtggatttc cgaagctgac agatgggtaa ggcgttgtg 49

<210> 310
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 310
 agtggatttc cgaagctgac agatgtggaa gaagaggctg gtcattgaggt caggagttcc 60

<210> 311
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 311
cgctcaggag gccttcaccc tctgctctgg gtaaaggaac tggaatatgc cttgaggggg 60

<210> 312
<211> 88
<212> DNA
<213> Homo sapiens

<400> 312
cctgagctgt ggtttgagg agccgtgtgt tggaagaaga tggcagatcc aggaatgatg 60
agtctttttg gcgaggatgg gaatatatt 88

<210> 313
<211> 79
<212> DNA
<213> Homo sapiens

<400> 313
cctgagctgt ggtttgagg agccgtgtgt tggaagaaga tcagaatgat gagtcttttt 60
ggcgaggatg ggaatatatt 79

<210> 314
<211> 78
<212> DNA
<213> Homo sapiens

<400> 314
cctgagctgt ggtttgagg agccgtgtgt tggaagaaga tggaatgatg agtctttttg 60
gcgaggatgg gaatatatt 78

<210> 315
<211> 60
<212> DNA
<213> Homo sapiens

<400> 315
cctgagctgt ggtttgagg agccgtgtgt tggaagaaga tggcgaggat gggaatatatt 60

<210> 316
<211> 41
<212> DNA
<213> Homo sapiens

<400> 316
taaagaaata tggaatgaca taaaatccag taaatcctat g 41

<210> 317
<211> 88
<212> DNA
<213> Homo sapiens

<400> 317
gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagtacaaca tccagaagaa 60
ggagattgtg gtgaaggagg acgaagtg 88

<210> 318

<211> 81
 <212> DNA
 <213> Homo sapiens

<400> 318
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagtacagaa gaaggagatt 60
 gtggtgaagg gagacgaagt g 81

<210> 319
 <211> 80
 <212> DNA
 <213> Homo sapiens

<400> 319
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagtacaacg aaggagattg 60
 tgggtgaagg agacgaagtg 80

<210> 320
 <211> 80
 <212> DNA
 <213> Homo sapiens

<400> 320
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga catccagaag aaggagattg 60
 tgggtgaagg agacgaagtg 80

<210> 321
 <211> 76
 <212> DNA
 <213> Homo sapiens

<400> 321
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagaagaagg agattgtggt 60
 gaagggagac gaagtg 76

<210> 322
 <211> 68
 <212> DNA
 <213> Homo sapiens

<400> 322
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagtattgtg gtgaaggag 60
 acgaagtg 68

<210> 323
 <211> 66
 <212> DNA
 <213> Homo sapiens

<400> 323
 gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagttgtggt gaagggagac 60
 gaagtg 66

<210> 324
 <211> 88
 <212> DNA

<213> Homo sapiens

<400> 324
gagggggggg aagatggcgg acgtgcttag cgtcctgcga cagtacaaca acatccagaa 60
gaaggagatt gtggtgaagg gagacgaa 88

<210> 325
<211> 56
<212> DNA
<213> Homo sapiens

<400> 325
agggcgaggc gacaagagaa gaaggaggca ggattgtggt gaaggagac gaagtg 56

<210> 326
<211> 88
<212> DNA
<213> Homo sapiens

<400> 326
ttagtatttt gaagttaata tcacaatgag ttcaggctta tggagccaag aaaaagtcac 60
ttcacctac tgggaagagc ggattttt 88

<210> 327
<211> 79
<212> DNA
<213> Homo sapiens

<400> 327
ttagtatttt gaagttaata tcacaatgag ttcaggctta tggaaagtca cttcaccta 60
ctgggaagag cggattttt 79

<210> 328
<211> 59
<212> DNA
<213> Homo sapiens

<400> 328
ttagtatttt gaagttaata tcacaagtca cttcaccta ctgggaagag cggattttt 59

<210> 329
<211> 47
<212> DNA
<213> Homo sapiens

<400> 329
ttagtatttt gaagttaata tcaccctact gggaagagcg gattttt 47

<210> 330
<211> 40
<212> DNA
<213> Homo sapiens

<400> 330
ttagtatttt gaagttccct actgggaaga gcggattttt 40

<210> 331
<211> 36

<212> DNA
 <213> Homo sapiens

 <400> 331
 taagtcactt caccctactg ggaagagcgg attttt 36

 <210> 332
 <211> 33
 <212> DNA
 <213> Homo sapiens

 <400> 332
 ttagtatttc cctactggga agagcggatt ttt 33

 <210> 333
 <211> 31
 <212> DNA
 <213> Homo sapiens

 <400> 333
 ttagtatattt gaagttgaag agcggatttt t 31

 <210> 334
 <211> 22
 <212> DNA
 <213> Homo sapiens

 <400> 334
 ttagtatattt gaacggattt tt 22

 <210> 335
 <211> 4
 <212> DNA
 <213> Homo sapiens

 <400> 335
 ttag 4

 <210> 336
 <211> 116
 <212> DNA
 <213> Homo sapiens

 <400> 336
 taatatcaca atgagttctc aggggcacac caggccccag gggaacacca ggccctgggt 60
 aagcatgcag tcccagggtg acatcagggtg ccaggaggaa aagtcacttc acccta 116

 <210> 337
 <211> 88
 <212> DNA
 <213> Homo sapiens

 <400> 337
 tccacacagc ctgtggcaca gacgtggagg gccactgagc cccgctaccc gccccacagc 60
 ctttcctacc cagtgcagat cgcccgga 88

 <210> 338
 <211> 87
 <212> DNA

<213> Homo sapiens

<400> 338

tccacacagc ctgtggcaca gacgtggagg gccactgagc cccgctaccg cccacagcc 60

tttcctaccc agtgcagatc gcccgga 87

<210> 339

<211> 81

<212> DNA

<213> Homo sapiens

<400> 339

tccacacagc ctgtggcaca gacgtggagg gccactgagc cccgccccac agcctttcct 60

accagtgca gatcgcccga a 81

<210> 340

<211> 76

<212> DNA

<213> Homo sapiens

<400> 340

tccacacagc ctgtggcaca gacgtggagg gccactgagc cccacagcct ttcctaccca 60

gtgcagatcg cccgga 76

<210> 341

<211> 19

<212> DNA

<213> Homo sapiens

<400> 341

tccacacaga tcgcccga 19

<210> 342

<211> 38

<212> DNA

<213> Homo sapiens

<400> 342

ccccaccac ctttcctacc cagtgcagat cgcccga 38

<210> 343

<211> 33

<212> DNA

<213> Homo sapiens

<400> 343

gcttattgcg gcctttctcc tgggagccag gct 33

<210> 344

<211> 88

<212> DNA

<213> Homo sapiens

<400> 344

tccacacagc ctgtggcaca gacgtggagg gccactgagc cccgctacca cccgccccac 60

agcctttcct accagtgca gatcgccc 88

<210> 345
 <211> 88
 <212> DNA
 <213> Homo sapiens

<400> 345
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg ggaatggcct gccttacgat 60
 gacagaaatg gagggaacat ccacctct 88

<210> 346
 <211> 84
 <212> DNA
 <213> Homo sapiens

<400> 346
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg ggcctgcct tacgatgaca 60
 gaaatggagg gaacatccac ctct 84

<210> 347
 <211> 83
 <212> DNA
 <213> Homo sapiens

<400> 347
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg ggcctgcctt acgatgacag 60
 aaatggaggg aacatccacc tct 83

<210> 348
 <211> 82
 <212> DNA
 <213> Homo sapiens

<400> 348
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg ggaatgggta cgatgacaga 60
 aatggagggg aacatccacct ct 82

<210> 349
 <211> 80
 <212> DNA
 <213> Homo sapiens

<400> 349
 tttctcttac aggcaaatgt tctgaaaaag actctgcagc ctgccttacg atgacagaaa 60
 tggaggggaa atccacctct 80

<210> 350
 <211> 73
 <212> DNA
 <213> Homo sapiens

<400> 350
 tttctcttac aggcaaatgt tctgaaaaag actctgcctt acgatgacag aaatggaggg 60
 aacatccacc tct 73

<210> 351
 <211> 64

<212> DNA
 <213> Homo sapiens

<400> 351
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg ggaatggagg gaacatccac 60
 ctct 64

<210> 352
 <211> 63
 <212> DNA
 <213> Homo sapiens

<400> 352
 tttctcttac aggcaaatgt tctgaaaaag acgatgacag aaatggaggg aacatccacc 60
 tct 63

<210> 353
 <211> 62
 <212> DNA
 <213> Homo sapiens

<400> 353
 tttctcttac aggcaaatgt tctgaaaaag actctgcaga aatggagggg acatccacct 60
 ct 62

<210> 354
 <211> 35
 <212> DNA
 <213> Homo sapiens

<400> 354
 tttctcttac aggcatggag ggaacatcca cctct 35

<210> 355
 <211> 34
 <212> DNA
 <213> Homo sapiens

<400> 355
 tacgatgaca gaaatggagg gaacatccac ctct 34

<210> 356
 <211> 84
 <212> DNA
 <213> Homo sapiens

<400> 356
 tttctcttac aggcaaatgt tctgaaaaag actctgcatg gagtaagtct tacgatgaca 60
 gaaatggagg gaacatccac ctct 84

<210> 357
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 357
 cggaggtttg ctgcaaacag agaaatttcg atgacagaaa tggagggaac atccacctct 60

<210> 358
 <211> 88
 <212> DNA
 <213> Homo sapiens

 <400> 358
 cgggaggcga gccgatgccg agctgctcca cgtccaccat gccgggcatg atctgcaaga 60
 acccagacct cgagtttgac tcgctaca 88

 <210> 359
 <211> 82
 <212> DNA
 <213> Homo sapiens

 <400> 359
 cgggaggcga gccgatgccg agctgctcca cgtcctcccc catgatctgc aggaaccag 60
 acctcgagtt tgactcgcta ca 82

 <210> 360
 <211> 80
 <212> DNA
 <213> Homo sapiens

 <400> 360
 cgggaggcga gccgatgccg agctgctcca cgtccgggca tgatctgcaa gaaccagac 60
 ctcgagtttg actcgctaca 80

 <210> 361
 <211> 78
 <212> DNA
 <213> Homo sapiens

 <400> 361
 cgggaggcga gccgatgccg agctgctcca cgtccaccat atctgcaaga acccagacct 60
 cgagtttgac tcgctaca 78

 <210> 362
 <211> 77
 <212> DNA
 <213> Homo sapiens

 <400> 362
 cgggaggcga gccgatgccg agctgctcca cgtccaccat gctgcaagaa ccagacctc 60
 gagtttgact cgctaca 77

 <210> 363
 <211> 76
 <212> DNA
 <213> Homo sapiens

 <400> 363
 cgggaggcga gccgatgccg agctgctcca cgtccatgat ctgcaagaac ccagacctcg 60
 agtttgactc gctaca 76

 <210> 364

<211> 70
 <212> DNA
 <213> Homo sapiens

<400> 364
 cgaggagcga gccgatgccg agctgctcca cgtccaccaa gaaccagac ctcgagtttg 60
 actcgctaca 70

<210> 365
 <211> 69
 <212> DNA
 <213> Homo sapiens

<400> 365
 cgaggagcga gccgatgccg agctgctcat gatctgcaag aaccagacc tcgagtttga 60
 ctcgctaca 69

<210> 366
 <211> 46
 <212> DNA
 <213> Homo sapiens

<400> 366
 aaggaagcac ccccggtctt aaagacctg agtttgactc gctaca 46

<210> 367
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 367
 agtgttgag gtcggcgccg gccccgggc atgatctgca agaaccaga cctcgagttt 60

<210> 368
 <211> 34
 <212> DNA
 <213> Homo sapiens

<400> 368
 aaggaagcac ccccggtatt aaacgaagat gact 34

<210> 369
 <211> 222
 <212> DNA
 <213> Homo sapiens

<400> 369
 gaggcgagcc gatgccgtg atctgggtca cggctgctcc agcttgagg agaggcggct 60
 ctcccgcgga ccctcctcgc gcgggcgccc ctgccattcc cggaacagg ggctcagccc 120
 cctttgttag tgctcgtatg tcttggcctg gggagcattt tggaggcagt gctaggggca 180
 gagaggtcct gtttccccca agtcttgatc tgcaagaacc ca 222

<210> 370
 <211> 88
 <212> DNA
 <213> Homo sapiens

<400> 370
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg aaactgctgc ccgccgcggg 60
 cccggcagga ggtaagggca gaagggaa 88

<210> 371
 <211> 83
 <212> DNA
 <213> Homo sapiens

<400> 371
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg gctgcccgcc gcgggcccgg 60
 caggaggtaa gggcagaagg gaa 83

<210> 372
 <211> 81
 <212> DNA
 <213> Homo sapiens

<400> 372
 tgcacgtcgg cccagccct gaggagccgg accgatgtgc tgcccgccgc gggcccggca 60
 ggaggttaagg gcagaaggga a 81

<210> 373
 <211> 75
 <212> DNA
 <213> Homo sapiens

<400> 373
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg aaactggccc ggcaggaggt 60
 aagggcagaa gggaa 75

<210> 374
 <211> 71
 <212> DNA
 <213> Homo sapiens

<400> 374
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg actcccggca ggaggttaagg 60
 gcagaaggga a 71

<210> 375
 <211> 69
 <212> DNA
 <213> Homo sapiens

<400> 375
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg acccggcagg aggttaagggc 60
 agaagggaa 69

<210> 376
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 376
 tgcacgtcgg cccagccct gaggagccgg accgatgcag gaggttaaggg cagaagggaa 60

<210> 377
 <211> 60
 <212> DNA
 <213> Homo sapiens

<400> 377
 tgcacgtcgg cccagccct gaggagccgg accgatgtgg aaagtaaggg cagaaggga 60

<210> 378
 <211> 58
 <212> DNA
 <213> Homo sapiens

<400> 378
 tgcacgtcgg cccagccct gaggagccgg accggcagga ggtaaggga gaaggga 58

<210> 379
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 379
 tgcacgtcgg cccagccct gaggagccgg aaggaggtaa gggcagaagg gaa 53

<210> 380
 <211> 49
 <212> DNA
 <213> Homo sapiens

<400> 380
 tgcacgtcgg cccgcgcgg gccggcagg aggtaaggga agaaggga 49

<210> 381
 <211> 43
 <212> DNA
 <213> Homo sapiens

<400> 381
 tgcacgtcgg cccagccgg caggaggtaa gggcagaagg gaa 43

<210> 382
 <211> 52
 <212> DNA
 <213> Homo sapiens

<400> 382
 gctcccgga gcgcgcacgt cccggagccc atgcctgcgg gtgattcctg cg 52

<210> 383
 <211> 88
 <212> DNA
 <213> Homo sapiens

<400> 383
 tcgcgaagtg gaatttggc agacaagcaa catggctcgg aaacgcgcgg ccggcgggga 60

gccgcgggga cgcgaactgc gcagccag 88

<210> 384

<211> 86
 <212> DNA
 <213> Homo sapiens

<400> 384
 tcgcgaagtg gaatttgccc agacaagcaa catggctcgg aggaacggcc ggcggggagc 60
 cgcggggacg cgaactgcg agccag 86

<210> 385
 <211> 76
 <212> DNA
 <213> Homo sapiens

<400> 385
 tcgcgaagtg gaatttgccc agacaagcaa catggctcgg aaaggggagc cgcggggacg 60
 cgaactgcg agccag 76

<210> 386
 <211> 73
 <212> DNA
 <213> Homo sapiens

<400> 386
 tcgcgaagtg gaatttgccc agacaagcaa catggccggc ggggagccgc ggggacgcga 60
 actgcgagc cag 73

<210> 387
 <211> 71
 <212> DNA
 <213> Homo sapiens

<400> 387
 tcgcgaagtg gaatttgccc agacaagcaa catggctcgg ggagccgcgg ggacgcgaac 60
 tgcgcagcca g 71

<210> 388
 <211> 69
 <212> DNA
 <213> Homo sapiens

<400> 388
 tcgcgaagtg gaatttgccc agacaagcaa catggcgggg agccgcgggg acgcgaactg 60
 cgcagccag 69

<210> 389
 <211> 53
 <212> DNA
 <213> Homo sapiens

<400> 389
 tcgcgaagtg gaatttgccc agacaagcaa catgttgca actgcgagc cag 53

<210> 390
 <211> 49
 <212> DNA
 <213> Homo sapiens

<400> 390
tcgCGAAGtg gaatttGccc agacaagcaa catggctcgg cgcagccag 49

<210> 391
<211> 12
<212> DNA
<213> Homo sapiens

<400> 391
tcgCGcagcc ag 12

<210> 392
<211> 31
<212> DNA
<213> Homo sapiens

<400> 392
tcgCGaAGtg gaatccaagg ccaagagcaa g 31

<210> 393
<211> 88
<212> DNA
<213> Homo sapiens

<400> 393
gtcGaccccg ctGcacagtc cggccggcgc catgaagtga gaagggggct gggggtcgcg 60
ctcGctagcg ggcGcggggg gtcttgaa 88

<210> 394
<211> 81
<212> DNA
<213> Homo sapiens

<400> 394
gtcGaccccg ctGcacagtc cggccggcgc catgaagtgg gctgggggtc gcgctcgcta 60
gcgggGcgcg ggggtcttga a 81

<210> 395
<211> 77
<212> DNA
<213> Homo sapiens

<400> 395
gtcGaccccg ctGcacagtc cggccggcgc catgaagctg ggggtcgcgc tcgctagcgg 60
gcGcgggggg tcttgaa 77

<210> 396
<211> 57
<212> DNA
<213> Homo sapiens

<400> 396
gtcGaccccg ctGcacagtc cggccggcgc tcgctagcgg gcGcgggggg tcttgaa 57

<210> 397
<211> 49
<212> DNA
<213> Homo sapiens

<400> 397
gtcgaccccc ctgcacagtc cggccggcgc catgaagtgg ggtcttgaa 49

<210> 398
<211> 43
<212> DNA
<213> Homo sapiens

<400> 398
ggcctctcgc tgaatattca tgatgggggc atcgggtgggc gcg 43

<210> 399
<211> 53
<212> DNA
<213> Homo sapiens

<400> 399
gtcgaccccc ctgcacagtc cggccggcgc catgaagggt gaaggggtgg gac 53

<210> 400
<211> 65
<212> DNA
<213> Homo sapiens

<400> 400
gtcgaccccc ctgcacagtc cggccggcgc catgaagtgc gtccgggggtg ggacgggggc 60
agccg 65

<210> 401
<211> 60
<212> DNA
<213> Homo sapiens

<400> 401
gtcgaccccc ctgcacagtc cggccggcgc catgaagtgg cagccgcagg gagcagcagt 60

<210> 402
<211> 51
<212> DNA
<213> Homo sapiens

<400> 402
atggtctcgt aatatagggtg gagcgagccc tcgagggggg tcttgaagat g 51

<210> 403
<211> 33
<212> DNA
<213> Homo sapiens

<400> 403
gtcgaccccc ctgctgagga cctgagggtt acc 33

<210> 404
<211> 42
<212> DNA
<213> Homo sapiens

<400> 404
ggcctctcgc tgaatattca gtcctgagg acctgagggt ta 42

<210> 405
 <211> 44
 <212> DNA
 <213> Homo sapiens

 <400> 405
 gtcgaccccg ctgcacagtc cggcccgctca cccttctctg ggct 44

 <210> 406
 <211> 40
 <212> DNA
 <213> Homo sapiens

 <400> 406
 gtcgaccccg ctgcacagtc cggccgctcg acgaccgggc 40

 <210> 407
 <211> 46
 <212> DNA
 <213> Homo sapiens

 <400> 407
 gtcgaccccg ctgcacagtc cggccggcga ccgggactg tggagg 46

 <210> 408
 <211> 39
 <212> DNA
 <213> Homo sapiens

 <400> 408
 gtcgaccccg ctgcacagtc cggccgggca ctgtggagg 39

 <210> 409
 <211> 19
 <212> DNA
 <213> Homo sapiens

 <400> 409
 gtcgaccag cagcccctg 19

 <210> 410
 <211> 19
 <212> DNA
 <213> Homo sapiens

 <400> 410
 gtcgaccccc ctgccgcca 19

 <210> 411
 <211> 43
 <212> DNA
 <213> Homo sapiens

 <400> 411
 ggcctctcgc tgaatattca tgagccgcca ggctcaacgt gga 43

 <210> 412
 <211> 18
 <212> DNA

<213> Homo sapiens
 <400> 412
 tatgtacgcc tccctggg 18

<210> 413
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400> 413
 tacagaagcg ggcaaagg 18

<210> 414
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400> 414
 tccgaagctg acagatgg 18

<210> 415
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400> 415
 tcaggttccg cccctacc 18

<210> 416
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400> 416
 ttagacttag gtaagtaa 18

<210> 417
 <211> 18
 <212> DNA
 <213> Homo sapiens
 <400> 417
 tagtttgtag ttctcccc 18

<210> 418
 <211> 17
 <212> DNA
 <213> Homo sapiens
 <400> 418
 tccggccggc gccatga 17

<210> 419
 <211> 16
 <212> DNA
 <213> Homo sapiens
 <400> 419
 tagcgagcgc gacccc 16

<210> 420
 <211> 18
 <212> DNA
 <213> Homo sapiens

 <400> 420
 taggcgcaa ggccatgt 18

<210> 421
 <211> 18
 <212> DNA
 <213> Homo sapiens

 <400> 421
 tggcccagg cgagttc 18

<210> 422
 <211> 18
 <212> DNA
 <213> Homo sapiens

 <400> 422
 tgaaggaca tcaccttt 18

<210> 423
 <211> 17
 <212> DNA
 <213> Homo sapiens

 <400> 423
 tctactgaat cttgagc 17

<210> 424
 <211> 16
 <212> DNA
 <213> Homo sapiens

 <400> 424
 tcggccacca tgtccc 16

<210> 425
 <211> 18
 <212> DNA
 <213> Homo sapiens

 <400> 425
 tccaggcagc tggagccc 18

<210> 426
 <211> 18
 <212> DNA
 <213> Homo sapiens

 <400> 426
 tctgaaaaag actctgca 18

<210> 427
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 427 tccatttctg tcatcgta	18
<210> 428 <211> 18 <212> DNA <213> Homo sapiens	
<400> 428 tgaaaatgac tgaatata	18
<210> 429 <211> 17 <212> DNA <213> Homo sapiens	
<400> 429 ttgcctacgc caccagc	17
<210> 430 <211> 18 <212> DNA <213> Homo sapiens	
<400> 430 tgcttagacg ctggattt	18
<210> 431 <211> 18 <212> DNA <213> Homo sapiens	
<400> 431 ttcggtgctt acctggtt	18
<210> 432 <211> 17 <212> DNA <213> Homo sapiens	
<400> 432 tcccagacat gacagcc	17
<210> 433 <211> 18 <212> DNA <213> Homo sapiens	
<400> 433 tccttttggt tctgctaa	18
<210> 434 <211> 18 <212> DNA <213> Homo sapiens	
<400> 434 ttgccgtccc aagcaatg	18

<210> 435
<211> 18
<212> DNA
<213> Homo sapiens

<400> 435
tgttcaatat cgtccggg 18

<210> 436
<211> 18
<212> DNA
<213> Homo sapiens

<400> 436
tgggccagag atggcggc 18

<210> 437
<211> 18
<212> DNA
<213> Homo sapiens

<400> 437
taaagccgcc gcctccgg 18

<210> 438
<211> 18
<212> DNA
<213> Homo sapiens

<400> 438
tgcccagaca agcaacat 18

<210> 439
<211> 18
<212> DNA
<213> Homo sapiens

<400> 439
tccccgcggc tccccgcc 18