

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(10) **DK/EP 2893022 T3**

(12) **Oversættelse af
europæisk patentskrift**

-
- (51) Int.Cl.: **C 12 N 15/62 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2020-07-27**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2020-05-13**
- (86) Europæisk ansøgning nr.: **13834893.3**
- (86) Europæisk indleveringsdag: **2013-09-04**
- (87) Den europæiske ansøgnings publiceringsdag: **2015-07-15**
- (86) International ansøgning nr.: **US2013058100**
- (87) Internationalt publikationsnr.: **WO2014039585**
- (30) Prioritet: **2012-09-04 US 201261696689 P** **2013-01-17 US 201361753763 P**
2013-05-01 US 201361818364 P
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA**
- (72) Opfinder: **BARBAS III, Carlos, F., 5416 Candlelight Drive, La Jolla, CA 92037, USA**
MERCER, Andrew, 19932 Westerly Avenue, Poolesville, MD 20837, USA
LAMB, Brian, M., 3144 Cowley Way Apt. 3, San Diego, CA 92117, USA
GAJ, Thomas, 215 Mosswood Way, San Francisco, CA 94080, USA
- (74) Fuldmægtig i Danmark: **NORDIC PATENT SERVICE A/S, Bredgade 30, 1260 København K, Danmark**
- (54) Benævnelse: **KIMÆRISKE POLYPEPTIDER MED MÅLRETTET BINDINGSSPECIFICITET**
- (56) Fremdragne publikationer:
WO-A1-2012/033462
WO-A2-2013/101877
US-A1- 2010 086 532
US-A1- 2011 301 073
US-A1- 2012 178 131
US-A1- 2012 270 273
C. MUSSOLINO ET AL: "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity", NUCLEIC ACIDS RESEARCH, vol. 39, no. 21, 3 August 2011 (2011-08-03), pages 9283-9293, XP055021128, ISSN: 0305-1048, DOI: 10.1093/nar/gkr597
JEFFREY C MILLER ET AL: "A TALE nuclease architecture for efficient genome editing Supplementary information", NATURE BIOTECHNOLOGY, vol. 29, no. 2, 22 December 2010 (2010-12-22), pages 143-148, XP055077697, ISSN: 1087-0156, DOI: 10.1038/nbt.1755
YU YANHUA ET AL: "Colonization of Rice Leaf Blades by an African Strain of Xanthomonas oryzae pv. oryzae Depends on a New TAL Effector That Induces the Rice Nodulin-3 Os11N3 Gene", MOLECULAR PLANT-MICROBE INTERACTIONS, vol. 24, no. 9, September 2011 (2011-09), pages 1102-1113, XP002752968, ISSN:

Fortsættes ...

0894-0282

OCHIAI HIROKAZU ET AL: "Genome sequence of Xanthomonas oryzae pv. oryzae suggests contribution of large numbers of effector genes and insertion sequences to its race diversity", JARQ, vol. 39, no. 4, October 2005 (2005-10), pages 275-287, XP002752969, ISSN: 0021-3551

MUKAIHARA TAKAFUMI ET AL: "Genetic screening of Hrp type III-related pathogenicity genes controlled by the HrpB transcriptional activator in Ralstonia solanacearum", MOLECULAR MICROBIOLOGY, vol. 54, no. 4, November 2004 (2004-11), pages 863-875, XP002752970, ISSN: 0950-382X

A. C. MERCER ET AL: "Chimeric TALE recombinases with programmable DNA sequence specificity", NUCLEIC ACIDS RESEARCH, vol. 40, no. 21, 1 November 2012 (2012-11-01), pages 11163-11172, XP055133292, ISSN: 0305-1048, DOI: 10.1093/nar/gks875

LAMB BRIAN M ET AL: "Directed evolution of the TALE N-terminal domain for recognition of all 5' bases", NUCLEIC ACIDS RESEARCH, vol. 41, no. 21, November 2013 (2013-11), pages 9779-9785, XP002752971, GAJ ET AL.: 'ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering' CELL vol. 31, no. 7, 09 May 2013, pages 397 - 405, XP055065263

DESCRIPTION

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0001] The present invention relates generally to the field of biotechnology, and more specifically to chimeric recombinases that recognize specific DNA sequences.

BACKGROUND INFORMATION

[0002] The ability of proteins to recognize DNA in a sequence-dependent manner is central to life, as a variety of protein domains have evolved to provide sequence-specific DNA recognition. DNA recognition by a select few of these domains is also the foundation for a wide variety of biotechnological applications. In particular, C₂H₂ zinc-finger proteins (ZFPs) were among the first DNA-binding proteins to be engineered to recognize user-defined DNA sequences and have been used with varying degrees of success for many applications, including transcriptional regulation, genome engineering and epigenetic modification. Modular assembly of ZFPs has facilitated these approaches. However, despite the advances and promise of ZFP technology, construction of specific, high-affinity ZFPs for certain sequences remains difficult and in select cases requires the use of time-consuming and labor-intensive selection systems not readily adopted by non-specialty laboratories.

[0003] Transcription activator-like effector (TALE) domains are a class of naturally occurring DNA-binding domains (DBDs) that represent a potential alternative to ZFP technology. TALEs, which are found in the plant pathogen *Xanthomonas*, contain a series of 33 to 35 amino acid repeats that function to selectively bind target DNA sequences. These repeats are identical with the exception of two adjacent repeat variable di-residues (RVDs) that confer DNA specificity by mediating binding to a single nucleotide. Arrays of over 30 repeats have been described that bind to DNA sites of similar numbers of base pairs (bps). Although there is inherent degeneracy in the binding of each RVD, recent reports have indicated that synthetic TALE proteins are specific enough to target single loci within the human genome.

[0004] The introduction of DNA double-strand breaks (DSBs) by chimeric nucleases, such as zinc-finger nucleases (ZFNs) can be used to knockout gene function or in the presence of exogenously added DNA drive cassette integration at the targeted loci. ZFNs have been extensively studied over the last decade and in some cases are approaching clinical use for gene therapy. Recently, a number of groups have explored the use of TALE DNA-binding domains fused to nucleases (TALENs) for targeted genome editing. Indeed, much of the work with ZFNs has been replicated with TALE nucleases, as TALENs may have advantages over ZFNs in regards to DNA-binding modularity. However, despite impressive research with ZFNs and

TALENs, questions remain about their safety and specificity. In particular, off-target cleavage events remain difficult to detect, as the most likely result of an off-target DSB is the introduction of small insertions or deletions. Additionally, repair of DSBs relies on cell machinery that varies with cell type.

[0005] An alternate approach for achieving targeted genomic modifications is the use of site-specific recombinases (SSRs). SSRs, such as the tyrosine recombinases Cre and Flp, are valuable molecular biology tools that are routinely used to manipulate chromosome structure inside cells. Because these enzymes rely on a number of complex protein-protein and protein-DNA interactions to coordinate catalysis, SSRs exhibit remarkable target site specificity. To date, however, altering the specificity of many SSRs has proven difficult. Serine recombinases of the resolvase/invertase type provide a versatile alternative to tyrosine recombinases for genome engineering. In nature, these enzymes function as multi-domain protein complexes that coordinate recombination in a highly modular manner. However, mutants of several serine recombinases have been identified that do not require accessory factors for recombination. Additionally, numerous studies have shown that the native DBDs of serine recombinases can be replaced with custom-designed ZFPs to generate chimeric zinc-finger recombinases (ZFRs). In principle, ZFRs capable of recognizing an extended number of sequences could be generated, however, the lack of zinc-finger domains capable of recognizing all possible DNA triplets limits the potential modular targeting capacity of these enzymes.

[0006] ZFRs are composed of an activated catalytic domain derived from the resolvase/invertase family of serine recombinases and a zinc-finger DNA-binding domain that can be custom-designed to recognize almost any DNA sequence (Figure 30A). ZFRs catalyze recombination between specific ZFR target sites that consist of two-inverted zinc-finger binding sites (ZFBS) flanking a central 20-bp core sequence recognized by the recombinase catalytic domain (Figure 30B). In contrast to zinc-finger nucleases (ZFNs) and TAL effector nucleases (TALENs), ZFRs function autonomously and can excise and integrate transgenes in human and mouse cells without activating the cellular DNA damage response pathway. However, as with conventional site-specific recombinases, applications of ZFRs have been restricted by sequence requirements imposed by the recombinase catalytic domain, which dictate that ZFR target sites contain a 20-bp core derived from a native serine resolvase/invertase recombination site.

[0007] Site-specific DNA recombination systems such as Cre-loxP, FLP-FRT and Δ C31-att have emerged as powerful tools for genetic engineering. The site-specific recombinases that promote these DNA rearrangements recognize short (30- to 40-bp) sequences and coordinate DNA cleavage, strand exchange and re-ligation by a mechanism that does not require DNA synthesis or a high-energy cofactor. This simplicity has allowed researchers to study gene function with extraordinary spatial and temporal sensitivity. However, the strict sequence requirements imposed by site-specific recombinases have limited their application to cells and organisms that contain artificially introduced recombination sites. In order to address this limitation, directed evolution has been used to alter the sequence specificity of several recombinases toward naturally occurring DNA sequences. Despite advances, the need for complex mutagenesis and selection strategies and the finding that re-engineered recombinase variants routinely exhibit relaxed substrate specificity have hindered the widespread adoption of this technology.

[0008] Mussolino et al 'A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity', Nucleic Acids Research, vol. 39, no.21, 2011, pages 9283-9293 discloses a systematic mutational analysis of RVD amino acid exchanges at positions 12 and 13 in relation to altered binding specificities of the mutated RVDs towards the four DNA bases. Miller et al 'A TALE nuclease architecture for efficient genome editing supplementary information', Nature Biotechnology, vol.29, no 2, 2010, pages 143-148 discloses TALEs specific for a 5' guanine.

[0009] Accordingly, there is a need for a more generalized method of catalyzing targeted and site-specific recombination of the endogenous genome, particularly for gene therapy, as well as for enzymes that can catalyze such targeted and site-specific recombination. This is particularly useful for gene therapy, but would have many other applications in molecular biology, including in gene cloning and use in modification of industrial organisms and agricultural plants and animals.

SUMMARY OF THE INVENTION

[0010] The invention is as defined in the claims.

[0011] In a first aspect of the invention, there is provided an isolated polypeptide comprising a *Xanthomonas* derived transcription activator-like effector (TALE) protein having an N-terminal domain (NTD) comprising an amino acid sequence in the N₁ hairpin as set forth in SEQ ID NO: 3 (VGKQWSGARAL) having one or more mutations selected from: Q is Y, Q is S, Q is R, W is R, W is G, W is deleted, S is R, S is H, S is A, S is N, and S is T.

[0012] In another aspect there is provided an isolated polypeptide comprising a *Ralstonia* derived transcription activator-like effector (TALE) protein having an N-terminal domain (NTD) comprising an amino acid sequence in the N₁ hairpin as set forth in SEQ ID NO: 8 (IVDIAR₁QR₂SGDLA) having one or more mutations selected from: R₁ is K, Q is Y, Q is S, Q is R, R₂ is W, R₂ is G, R₂ is deleted, S is R, S is H, S is A, S is N, and S is T.

[0013] In another aspect there is provided a method of generating a transcription activator-like effector (TALE) protein N-terminal domain (NTD), comprising:

1. a) randomizing an amino acid sequence of the NTD by mutating one or more amino acid residues within the NTD, wherein the amino acid sequence is SEQ ID NO: 14 (VGKXXXGAR) or SEQ ID NO: 15 (VDIAXXXGDLA); and
2. b) selecting for the randomized TALE protein NTD of (a), wherein the TALE protein NTD specifically binds to a desired nucleotide or exhibits enhanced activity.

[0014] In another aspect there is provided a method for site-specific recombination comprising:

1. (a) providing a DNA sequence comprising at least two binding sites for specifically

- interacting with a polypeptide of the invention; and
2. (b) reacting the DNA sequence with the polypeptide, wherein the polypeptide catalyzes a site-specific recombination event in which both strands of the DNA sequence are cleaved between the two sites specifically interacting with the polypeptide.

[0015] In another aspect of the invention, there is provided composition comprising a nucleic acid molecule encoding the polypeptide of the invention, for use in gene therapy.

[0016] In another aspect there is provided a pharmaceutical composition comprising:

1. (a) a polypeptide of the invention, or a nucleic acid encoding the polypeptide; and
2. (b) a pharmaceutically acceptable carrier.

[0017] Disclosed herein are targeted chimeric polypeptides, including compositions thereof, expression vectors, and methods of use thereof, for the generation of transgenic cells, tissues, plants, and animals. The compositions, vectors, and methods of the present invention are also useful in gene therapy techniques.

[0018] In one aspect, the invention provides a chimeric polypeptide. The polypeptide includes: a) a recombinase, nuclease or transcription factor, or fragment thereof; and b) a transcription activator-like effector (TALE) protein. In various embodiments, the TALE protein is truncated and includes a C-terminal or N-terminal truncation. In embodiments, the TALE protein is AvrXa7, Tal1c, and PthXo1. In embodiments, the TALE protein includes all or a portion an amino acid sequence as set forth in SEQ ID NO: 2. In some embodiments, the TALE protein is truncated between amino acid residues 27 and 268, 92 and 134, 120 and 129, 74 and 147, or 87 and 120 of SEQ ID NO: 2. In some embodiments, the TALE protein is truncated at amino acid residue 28, 74, 87, 92, 95, 120, 124, 128, 129, 147 and 150 of SEQ ID NO: 2.

[0019] In another aspect, the invention provides a method of generating a transcription activator-like effector (TALE) protein binding domain which specifically binds a desired nucleotide. The method includes a) randomizing the amino acid sequence of the TALE protein binding domain by mutating an amino acid residue within a variable di-residue (RVD), or within 1 to 2 amino acid residues N-terminal or C-terminal of the RVD; and b) selecting for the randomized TALE protein binding domain of (a), wherein the TALE protein binding domain specifically binds to the desired nucleotide.

[0020] In another aspect, the invention provides an isolated polypeptide comprising a *Xanthomonas* derived transcription activator-like effector (TALE) protein, the TALE protein having an N-terminal domain (NTD) comprising an amino acid sequence as set forth in SEQ ID NO: 3 (VGKQWSGARAL) having one or more mutations or deletions selected from: Q is Y, Q is S, Q is R, W is R, W is G, W is deleted, S is R, S is H, S is A, S is N, and S is T.

[0021] In another aspect, the invention provides an isolated polypeptide including a *Ralstonia* derived transcription activator-like effector (TALE) protein, the TALE protein having an N-terminal domain (NTD) including an amino acid sequence as set forth in SEQ ID NO: 8 (IVDIAR₁QR₂SGDLA) having one or more mutations or deletions selected from: R₁ is K, Q is Y, Q is S, Q is R, R₂ is W, R₂ is G, R₂ is deleted, S is R, S is H, S is A, S is N, and S is T.

[0022] In another embodiment, the invention provides a method of generating a transcription activator-like effector (TALE) protein N-terminal domain (NTD). The method includes: a) randomizing an amino acid sequence of the NTD by mutating or deleting one or more amino acid residues within the NTD, wherein the amino acid sequence is SEQ ID NO: 14 (VGKXXXGAR) or SEQ ID NO: 15 (VDIAXXXGDLA); and b) selecting for the randomized TALE protein NTD of (a), wherein the TALE protein NTD specifically binds to a desired nucleotide or exhibits enhanced activity.

[0023] Also disclosed herein are chimeric proteins including a serine recombinase and one or more zinc finger binding domains, methods of generating ZFRs, compositions thereof, expression vectors, and methods of use thereof, for the generation of transgenic cells, tissues, plants, and animals. The compositions, vectors, and methods of the present invention are also useful in gene therapy techniques.

[0024] In one aspect, the invention provides a method of generating a plurality of zinc finger recombinase (ZFRs) proteins having catalytic specificity greater than the corresponding wild type recombinase. The method includes performing random mutagenesis on a recombinase catalytic domain at positions equivalent to Gin Ile120, Thr123, Leu127, Ile136 and Gly137 or a combination thereof, mutating the DNA at positions 2 and 3 for each amino acid; fusing the recombinase catalytic domain with a plurality of zinc finger binding domains to form ZFRs, and enriching for ZFRs having catalytic specificity greater than the corresponding wild type recombinase. In embodiments the ZFRs have increased catalytic activity on DNA targets selected from GC, GT, CA, TT and AC. In one embodiment, the recombinase catalytic domain is mutagenized at Ile136 and/or Gly137.

[0025] In various aspects, the chimeric polypeptides described herein include a recombinase catalytic domain derived from or randomly mutagenized as disclosed herein from: a) Tn3, also known as EcoTn3; Hin, also known as StyHin; Gin, also known as MuGin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsoISC1904a; SsoISC1904b; SsoISC1913; Aam606; MjaM0014; Pab; HpyIS607; MtuIS_Y349; MtuRv2792c; MtuRv2979c; MtuRv3828c; MtuRv0921; MceRv0921; TnpX; TndX; WwK; lactococcal phage TP901-1 serine recombinase; *S. pyogenes* phage ϕ 370.1 serine recombinase; *S. pyogenes* phage ϕ FC1 serine recombinase; *Listeria* phage A118 serine recombinase; *S. coelicolor* chromosome SC3C8.24 serine recombinase; *S. coelicolor* chromosome SC2E1.37 serine recombinase; *S. coelicolor* chromosome SCD78.04c serine recombinase; *S. coelicolor* chromosome SC8F4.15c serine recombinase; *S. coelicolor* chromosome SCD12A.23 serine recombinase; *S. coelicolor* chromosome SCH10.38c serine recombinase; *S. coelicolor* chromosome SCC88.14 serine recombinase; *Streptomyces* phage

φC31 serine recombinase; *Streptomyces* phage R4 serine recombinase; *Bacillus* phage φ105 serine recombinase; *Bacillus* phage SPBc2 serine recombinase; *Bacillus* prophage SKIN serine recombinase; *S. aureus ccrA* serine recombinase; *S. aureus ccrB* serine recombinase; *M. tuberculosis* phage Bxb1 serine recombinase; *M. tuberculosis* prophage φPV1 serine recombinase; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; and Cac 1954; or b) muteins of a).

[0026] In yet another aspect, the invention provides an isolated nucleic acid molecule encoding the chimeric polypeptide of the invention.

[0027] In yet another aspect, the invention provides an expression cassette including the nucleic acid molecule the chimeric polypeptide of the invention.

[0028] In yet another aspect, the invention provides a vector including the expression cassette of the invention.

[0029] In yet another aspect, the invention provides an isolated host cell containing the vector of the invention.

[0030] In yet another aspect, the invention provides a method for site-specific integration into a DNA sequence. The method includes contacting the DNA sequence with a chimeric polypeptide of the present invention, wherein the chimeric polypeptide catalyzes site-specific integration.

[0031] In yet another example, the invention provides a method for gene therapy. The method includes administering to a subject a composition comprising a nucleic acid molecule encoding the chimeric polypeptide described herein, wherein upon expression of the nucleic acid molecule, a gene present in the genome of the subject is specifically removed or inactivated.

[0032] In yet another aspect, the invention provides a pharmaceutical composition. The composition includes the chimeric polypeptide described herein; and a pharmaceutically acceptable carrier. In another aspect, the composition includes a nucleic acid molecule encoding the chimeric polypeptide of the invention; and a pharmaceutically acceptable carrier.

[0033] In yet another example, the disclosure provides a transgenic organism produced by recombination catalyzed by the chimeric polypeptide of the present disclosure.

[0034] In yet another example, the disclosure provides a method for gene therapy. The method includes administering to a subject a cell comprising a nucleic acid molecule having the DNA sequence generated by the method of site-specific integration described herein.

[0035] In another aspect, the invention provides an isolated nucleic acid molecule encoding the chimeric protein of the invention.

[0036] In another aspect, the invention provides a method for site-specific recombination. The method includes: a) providing a DNA sequence comprising at least two binding sites for specifically interacting with the chimeric protein of the invention; and b) reacting the DNA

sequence with the chimeric protein, wherein the chimeric protein catalyzes a site-specific recombination event in which both strands of the DNA sequence are cleaved between the two sites specifically interacting with the chimeric protein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0037] Figure 1 is a series of graphical and diagrammatic representations regarding TALER fusion orientation and activity. A) Cartoon illustrating the split β -lactamase system used to evaluate TALER activity. B) Schematic showing the fusion orientation of each TALER and its corresponding target site (1=SEQ ID NO: 288; 2=SEQ ID NO: 289; 3=SEQ ID NO: 290). C) Activity of each designed TALER fusion against its intended DNA target. Recombination was normalized to background (vector only control). D) Gin-Avr activity against cognate (Avr-20G) and non-cognate (Avr-20T, Avr-20GG, PthXo1-20G) DNA targets. Error bars indicate standard deviation (s.d.) (n = 3).

Figure 2 is a series of graphical and diagrammatic representations regarding recombination profiles of selected TALER truncations. A) Schematic illustrating the design of the 20-member TALER truncation library. B) Activity of selected TALER variants against DNA targets containing core sequences of increasing length (14, 20, 26, 32 and 44-bp). C) Gin-AvrXa7 Δ 120 activity against a diverse panel of substrates containing non-cognate cores sequences or core sites of increasing length. Error bars indicate s.d. (n = 3).

Figure 3 is a series of graphical representations regarding TALER variants selected from incremental truncation library. A) Frequency of selected TALER truncation variants. After 3 rounds of selection, incrementally truncated Gin-AvrXa7 variants were isolated and DNA sequencing was used to determine truncation length. B) Activity of incrementally truncated TALER variants (between Δ 92 and Δ 134 in length) against the Avr-32G DNA target. For reference, the shortest (Δ 145) and longest (Δ 74) truncation variants, as well as Δ 87 were included. C) Activity of Gin-Avr Δ 74, Gin-Avr Δ 128 and Gin-Avr Δ 145 against a diverse panel of cognate and non-cognate DNA targets. Error bars indicate s.d. (n = 3).

Figure 4 is a series graphical representations regarding activity of synthetic TALERs. A) Activity of synthetic Gin-Avr15 Δ 128, Gin-Avr15 Δ 120 and Gin-Pth15 Δ 120 variants against the DNA targets Avr-32G or Pth-32G. B) Activity of synthetic TALERs with DBDs between 15 and 20 repeats in length based on Gin-Avr Δ 120 against Avr-32G and Avr-32T. Error bars indicate s.d. (n = 3).

Figure 5 is a series of graphical representations regarding TALER activity in mammalian cells. (A, B) Fold-reduction of luciferase expression in HEK293T cells co-transfected with (A) TALER or ZFR expression vectors (Gin-Avr Δ 120 and GinC4) in the presence of reporter plasmid (Avr-32G, Avr-44G and C4-20G) or (B) TALER and ZFR expression vector in combination (Gin-Avr Δ 120 + GinC4) with reporter plasmid (Avr-G-ZF). Error bars indicate s.d. (n = 3).

Figure 6 is a diagrammatic representation of location of primers for N-terminal designed truncations of AvrXa7 (SEQ ID NO: 1 DNA sequence; SEQ ID NO: 2 amino acid sequence). Star denotes the location of Δ 120 fusion point.

Figure 7 is a diagrammatic representation of a comparison of native wild-type and synthetic RDV domains for the AvrXa7 target sequence (SEQ ID NOs: 16-18).

Figure 8 is a diagrammatic representation of TALE and TALER amino acid sequences of AvrXa7 protein (SEQ ID NO: 19).

Figure 9 is a diagrammatic representation of construct AvrXa7 DNA sequence (SEQ ID NO: 20).

Figure 10. is a diagrammatic representation of construct Gin-Avr Δ 74 amino acid sequence (SEQ ID NO: 21).

Figure 11 is a diagrammatic representation of construct Gin-Avr Δ 87 amino acid sequence (SEQ ID NO: 22).

Figure 12 is a diagrammatic representation of construct Gin-Avr Δ 120 amino acid sequence (SEQ ID NO: 23).

Figure 13 is a diagrammatic representation of construct Gin-Avr Δ 120* amino acid sequence (SEQ ID NO: 24).

Figure 14 is a diagrammatic representation of construct Gin-Avr Δ 147 amino acid sequence (SEQ ID NO: 25).

Figure 15 is a diagrammatic representation of construct GinAvr15 Δ 128-synthetic protein amino acid sequence (SEQ ID NO: 26).

Figure 16 is a diagrammatic representation of construct Gin-Avr15 Δ 128-synthetic protein DNA sequence (SEQ ID NO: 27).

Figure 17 is a diagrammatic representation of construct GinAvr15 Δ 128-synthetic protein amino acid sequence (SEQ ID NO: 28).

Figure 18 is a series of pictorial and graphical representations pertaining to the specificity of the TALE N-terminal domain. A) Illustration of a TALE (SEQ ID NO: 29) bound to its target DNA. B) Structural analysis suggests contact of the 5' T by W232 of the N-1 hairpin (N-0 - SEQ ID NO: 30; N-1 - SEQ ID NO: 31; and RVD - SEQ ID NO: 32). This hairpin shares significant sequence homology with RVD hairpins. C-F) Analyses of NT-T (wt) NTD in the context of C) AvrXa7 TALE-R, D) AvrXa7 TALE-TF, E) AvrXa7 MBPTALE, and F) a CCR5 targeting TALEN. (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$ compared to 5'T).

Figure 19 is a series of graphical and diagrammatic representations pertaining to recombinase variants. A-C) Activities of recombinase selection variants against substrates with A) 5' G, B) 5' A, and C) 5' C. Figure 18D is an alignment of optimized TALE NTDs SEQ ID NOs: 33-36), illustrating sequence differences in the N-1 hairpin. E) Comprehensive comparison of optimized NTD activities in the context of MBP-TALE AvrXa7. (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, compared to wild type and 5'A/G/C).

Figure 20 is a series of diagrammatic and graphical representations of analysis of selected NTDs in the context of TALE-TFs. A) Illustration of 5xAvr promoter region (SEQ ID NO: 37) on the luciferase reporter plasmid used for transcription activation experiments. B) Relative luciferase

activation of substrates with indicated 5' residues by TALE-TFs with NT-T, NT-G, NT-âN, and NT-âN domains. (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, compared to NT-T and respective 5'A/G/C/T).

Figure 21 is a series of diagrammatic and graphical representations of design and activity of TALEN pairs with wild-type and evolved NTD's with varying 5' bases. A) The CCR5 gene (SEQ ID NOs: 38-39) expanded to highlight the target site (SEQ ID NOs: 40-47) for induction of the H32 mutation. B) Gene editing efficiency of the wild type (NT-T) TALEN, TALENs with domains optimized for non-T 5' residues, and dHax3 NTD. C) Fold enhancement of the TALEN pairs with optimized NTD vs. TALENs with 5' T specificity. The activity of each NTD is shown on each TALEN pair substrate.

Figure 22 is a diagrammatic representation showing alignment of N- and C-terminal domains SEQ ID NOs: 48-53).

Figure 23 is a schematic representation illustrating TALE-Recombinase selection protocol. A library of NTD was cloned into Avr15 TALE-R using Not1/Stu1 restriction enzymes and complementary ligation. Active TALE-R's result in more frequent recombination events that can be selected and amplified with antibiotics (carbenecillin). The resulting output plasmid was the digested Not1/Xba1 and ligated into the TALE-R backbone vector for further selection and amplification.

Figure 24 is a diagrammatic representation of a summary of variant populations discovered from library selections (Library XXXSGAR (SEQ ID NO: 39) and Library KXXGAR (SEQ ID NO: 291)).

Figure 25 is a diagrammatic representation showing alignment of NT-G (SEQ ID NO: 54) with NTD-Brg11 (SEQ ID NO: 55), a *Ralstonia* TALE domain. Alignment indicates Brg11 could exhibit specificity for 5' G bases.

Figure 26 is a series of graphical representations of relative binding affinity of MBP-TALE proteins to target 5' A/G/C/T Avr15 hairpin oligonucleotides as assayed by ELISA. Protein concentrations were ~ 75nM and plates were developed for 120 minutes.

Figure 27 is a series of pictorial and graphical representations of a cell assay of PCR amplified CCR5 after TALEN editing with % indels and indel populations shown on the right.

Figure 28 is a diagrammatic representation showing alignment indel sequencing of selected TALEN experiments from Figure 27 (SEQ ID NOs: 292-332 from top to bottom).

Figure 29 is a graphical representation of a comparison of the activity of two separate Goldy TALE-Transcription factor architectures, each targeting identical 5x AvrXa7 promoters varying only in the 5' residue.

Figure 30 is a series of diagrammatic representations relating to the structure of the zinc-finger recombinase dimer bound to DNA. A) Each zinc-finger recombinase (ZFR) monomer (blue or orange) consists of an activated serine recombinase catalytic domain linked to a custom-designed zinc-finger DNA-binding domain. Model was generated from crystal structures of the $\gamma\delta$ resolvase and Aart zinc-finger protein (PDB IDs: 1GDT and 2113, respectively). B) Cartoon of the

ZFR dimer bound to DNA (SEQ ID NOs: 333-334). ZFR target sites consist of two-inverted zinc-finger binding sites (ZFBS) flanking a central 20-bp core sequence recognized by the ZFR catalytic domain. Zinc-finger proteins (ZFPs) can be designed to recognize 'left' or 'right' half-sites (blue and orange boxes, respectively). Abbreviations are as follows: N indicates A, T, C, or G; R indicates G or A; and Y indicates C or T.

Figure 31 is a series of graphical and diagrammatic representations of specificity of the Gin recombinase catalytic domain. A-D) Recombination was measured on DNA targets that contained (A, SEQ ID NO: 335) each possible two-base combination at the dinucleotide core, (B, SEQ ID NO: 336) each possible two-base combination at positions 3 and 2, (C, SEQ ID NO: 337) each possible single-base substitution at positions 6, 5, and 4, and (D, SEQ ID NO: 338) each possible single-base substitution at positions 10, 9, 8, and 7. Substituted bases are boxed above each panel. Recombination was evaluated by split gene reassembly and measured as the ratio of carbenicillin-resistant to chloramphenicol-resistant transformants (Materials and Methods). Error bars indicate standard deviation ($n = 3$). (E) Interactions between the $\gamma\delta$ resolvase dimer and DNA at (left) the dinucleotide core, (middle) positions 6, 5, and 4, and (right) positions 10, 9, 8, and 7 (PDB ID: 1GDT). Interacting residues are shown as purple sticks. Bases are colored as follows: A, yellow; T, blue; C, brown; and G, pink.

Figure 32 is a series of graphical and diagrammatic representations of re-engineering Gin recombinase catalytic specificity. A) The canonical 20-bp core recognized by the Gin catalytic domain. Positions 3 and 2 are boxed (SEQ ID NO: 339). B) (Top) Structure of the $\gamma\delta$ resolvase in complex with DNA (PDB ID: 1GDT). Arm region residues selected for mutagenesis are shown as purple sticks. (Bottom) Sequence alignment of the $\gamma\delta$ resolvase (SEQ ID NO: 341) and Gin recombinase (SEQ ID NO: 342) catalytic domains. Conserved residues are shaded orange. Black arrows indicate arm region positions selected for mutagenesis. C) Schematic representation of the split gene reassembly selection system. Expression of active ZFR variants leads to restoration of the β -lactamase reading frame and host-cell resistance to ampicillin. Solid lines indicate the locations and identity of the ZFR target sites. Positions 3 and 2 are underlined (SEQ ID NO: 340). D) Selection of Gin mutants that recombine core sites containing GC, GT, CA, TT, and AC base combinations at positions 3 and 2. Asterisks indicate selection steps in which incubation time was decreased from 16 hr to 6 hr (Materials and Methods, Example 5). E) Recombination specificity of the selected catalytic domains (β , γ , δ , ϵ , and ζ , wild-type Gin indicated by α) for each possible two-base combination at positions 3 and 2. Intended DNA targets are underlined. Recombination was determined by split gene reassembly and performed in triplicate.

Figure 33 is a series of graphical and diagrammatic representations illustrating the ability of ZFRs to recombine user-defined sequences in mammalian cells. A) Schematic representation of the luciferase reporter system used to evaluate ZFR activity in mammalian cells. ZFR target sites flank an SV40 promoter that drives luciferase expression. Solid lines denote the 44-bp consensus target sequence used to identify potential ZFR target sites. Underlined bases indicate zinc-finger targets and positions 3 and 2 (SEQ ID NO: 343). B) Fold-reduction of luciferase expression in HEK293T cells co-transfected with designed ZFR pairs and their cognate reporter plasmid. Fold-reduction was normalized to transfection with empty vector and reporter plasmid. The sequence identity and chromosomal location of each ZFR target site (SEQ ID NOs: 344-362 top to bottom)

and the catalytic domain composition of each ZFR pair are shown. Underlined bases indicate positions 3 and 2. Standard errors were calculated from three independent experiments. ZFR amino acid sequences are provided in Table 2. C) Specificity of ZFR pairs. Fold-reduction of luciferase expression was measured for ZFR pairs 1 through 9 and GinC4 for each non-cognate reporter plasmid. Recombination was normalized to the fold-reduction of each ZFR pair with its cognate reporter plasmid. Assays were performed in triplicate.

Figure 34 is a series of graphical and diagrammatic representations illustrating ZFRs ability to target integration into the human genome. A) Schematic representation of the donor plasmid (top) and the genomic loci targeted by ZFRs 1 (SEQ ID NO: 363), 2 (SEQ ID NO: 364), and 3 (SEQ ID NO: 365). Open boxes indicate neighboring exons. Arrows indicate transcript direction. The sequence and location of each ZFR target are shown. Underlined bases indicate zinc-finger targets and positions 3 and 2. B) Efficiency of ZFR-mediated integration. Data were normalized to data from cells transfected with donor plasmid only. Error bars indicate standard deviation ($n = 3$). C) PCR analysis of ZFR-mediated integration. PCR primer combinations amplified (top) unmodified locus or integrated plasmid in (middle) the forward or (bottom) the reverse orientation. D) Representative chromatograms of PCR-amplified integrated donor for ZFRs 1 (SEQ ID NO: 366) and 3 (SEQ ID NO: 367). Arrows indicate sequencing primer orientation. Shaded boxes denote genomic target sequences.

Figure 35 is a diagrammatic representation of recombinase DNA-binding residues are located outside the dimer interface. The $\gamma\delta$ resolvase in complex with target DNA. Catalytic domain dimer is colored cyan. DNA is colored grey. Arm region residues are shown as red sticks. Residues at the dimer interface are shown as purple sticks (PDB ID: 1GDT).

Figure 36 is a diagrammatic representation of sequence analysis of selected recombinases. Pie charts showing the percentage of amino acid substitutions at each targeted arm position. After the 4th round of selection, >20 clones were sequenced from each library. Sequence analysis of clones that recombine TT are described elsewhere(1).

Figure 37 is a table showing core specificity of isolated catalytic domains. After 4 rounds of selection, the ability of selected catalytic domains to recombine core sequences with substitutions at positions 3 and 2 was evaluated. Assigned DNA targets are underlined. Recombinase mutations are shown. Asterisks indicate catalytic domains selected for further analysis. Wild-type base combination at positions 3 and 2 is CC. Recombination was determined by split gene reassembly(2) and performed in triplicate. Catalytic domains that recombine TT substitutions are described elsewhere(1).

Figure 38 is a series of graphical representations of position specificity of selected catalytic domains. Recombination assays between the α , β , γ , δ and ζ catalytic domains and symmetrically substituted target sites. Recombination was measured on a library DNA targets that contained (A (SEQ ID NO: 368)) >4,000 random strong base (S: G or C) substitutions at positions 6, 5 and 4 and (B (SEQ ID NO: 369)) >10⁶ (of a possible 4.29×10^9) unique base combinations at positions 10, 9, 8 and 7 (N: A, T, C or G). Recombination was measured by split gene reassembly(2) ($n = 3$).

Figure 39 is a series of graphical representations of ZFR homodimer activity. HEK293T cells were

co-transfected with 150 ng ZFR-L or 150 ng ZFR-R with 2.5 ng of corresponding pGL3 ZFR reporter plasmid. Recombination was normalized to co-transfection with 150 ng ZFR-L and 150 ng ZFR-R with 2.5 ng pGL3 ZFR reporter plasmid.

Figure 40 is a series of pictorial representations depicting clonal analysis of ZFR-modified cells. PCR primer combinations amplified either unmodified genomic target or integrated plasmid in the forward or reverse orientation.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The present provides the first disclosure of a TALE recombinase (TALER). Using a library of incrementally truncated TALE domains, optimized TALER architecture that can be used to recombine DNA in bacterial and mammalian cells was identified. Any customized TALE repeat array can be inserted into the TALER architecture described herein, thus dramatically expanding the targeting capacity of engineered recombinases for applications in biotechnology and medicine.

[0039] Transcription activator-like effector (TALE) proteins can be designed to bind virtually any DNA sequence. General guidelines for design of TALE DNA-binding domains suggest that the 5'-most base of the DNA sequence bound by the TALE (the N_0 base) should be a thymine. The N_0 requirement was quantified by analysis of the activities of TALE transcription factors (TALE-TF), TALE recombinases (TALE-R) and TALE nucleases (TALENs) with each DNA base at this position. In the absence of a 5' T, decreases in TALE activity up to >1000-fold in TALE-TF activity, up to 100-fold in TALE-R activity and up to 10-fold reduction in TALEN activity compared with target sequences containing a 5' T was observed. To develop TALE architectures that recognize all possible N_0 bases, structure-guided library design coupled with TALE-R activity selections were used to evolve novel TALE N-terminal domains to accommodate any N_0 base. A G-selective domain and broadly reactive domains were isolated and characterized. The engineered TALE domains selected in the TALE-R format demonstrated modularity and were active in TALE-TF and TALEN architectures. Evolved N-terminal domains provide effective and unconstrained TALE-based targeting of any DNA sequence as TALE binding proteins and designer enzymes.

[0040] Additionally, in order to address sequence requirement limitations, a knowledge-base approach was described for re-engineering serine recombinase catalytic specificity. This strategy, which was based on the saturation mutagenesis of specificity-determining DNA-binding residues, was used to generate recombinase variants that showed a >10,000-fold shift in specificity. Importantly, this approach focused exclusively on amino acid residues located outside the recombinase dimer interface (Figure 35). As a result, it was determined that re-engineered catalytic domains could associate to form ZFR heterodimers and that these designed ZFR pairs recombine pre-determined DNA sequences with exceptional specificity. Together, these results led us to hypothesize that an expanded catalog of specialized catalytic domains developed by this method could be used to generate ZFRs with custom specificity. Here, a combination of substrate

specificity analysis and directed evolution is used to develop a diverse collection of Gin recombinase catalytic domains that are capable of recognizing an estimated 4×10^8 unique 20-bp core sequences. It is shown that ZFRs assembled from these re-engineered catalytic domains recombine user-defined sequences with high specificity and integrate DNA into targeted endogenous loci in human cells. These results demonstrate the potential of ZFR technology for a wide variety of applications, including genome engineering and gene therapy.

[0041] Before the present compositions and methods are described, it is to be understood that this invention is not limited to the particular compositions, methods, and experimental conditions described, as such devices, methods, and conditions may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only in the appended claims.

[0042] As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, references to "the composition" or "the method" includes one or more compositions and methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

[0043] Unless defined otherwise, 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 any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are now described.

[0044] "Recombinases" are a family of enzymes that mediate site-specific recombination between specific DNA sequences recognized by the recombinase (Esposito, D., and Scocca, J. J., *Nucleic Acids Research* 25, 3605-3614 (1997); Nunes-Duby, S. E., et al., *Nucleic Acids Research* 26, 391-406 (1998); Stark, W. M., et al., *Trends in Genetics* 8, 432-439 (1992)).

[0045] As used herein, the term "chimeric TALE recombinase" includes without limitation recombinases having a TALE domain derived from a naturally-occurring TALE protein or a synthetically derived TALE protein or domain with sequence-specific binding activity.

[0046] As used herein, the term "chimeric zinc finger recombinase" includes without limitation recombinases having a zinc finger binding domain derived from a naturally-occurring zinc finger DNA binding protein or a synthetically derived zinc finger binding protein or domain with sequence-specific binding activity.

[0047] As used herein, the term "zinc finger," "zinc finger nucleotide binding domain," or similar terminology refers both to naturally occurring and artificially produced zinc fingers. Such zinc fingers can have various framework structures, such as, but not limited to, C2H2, C4, H4, H3C, C3X, H3X, C2X2, and H2X2, where X is a zinc ligating amino acid. In these framework structures, as is conventional in the recitation of zinc finger structures, "C" represents a cysteine residue and

"H" represents a histidine residue. Zinc fingers having the framework C₂H₂ include, but are not limited to, zinc fingers described, for example, in International Publication Number WO2008/006028 to Barbas et al., United States Patent No. 7,101,972 to Barbas, United States Patent No. 7,067,617 to Barbas et al., United States Patent No. 6,790,941 to Barbas et al., United States Patent No. 6,610,512 to Barbas, United States Patent No. 6,242,568 to Barbas et al., United States Patent No. 6,140,466 to Barbas et al., United States Patent No. 6,140,081 to Barbas, United States Patent Application Publication No. 20060223757 by Barbas, United States Patent Application Publication No. 20060211846 by Barbas et al., United States Patent Application Publication No. 20060078880 by Barbas et al., United States Patent Application Publication No. 20050148075 by Barbas, United States Patent Application Publication No. 20050084885 by Barbas et al., United States Patent Application Publication No. 20040224385 by Barbas et al., United States Patent Application Publication No. 20030059767 by Barbas et al., and United States Patent Application Publication No. 20020165356 by Barbas et al.. Other zinc fingers are described in: U.S. Patent No. 7,067,317 to Rebar et al.; U.S. Patent No. 7,030,215 to Liu et al.; U.S. Patent No. 7,026,462 to Rebar et al.; U.S. Patent No. 7,013,219 to Case et al.; U.S. Patent No. 6,979,539 to Cox III et al.; U.S. Patent No. 6,933,113 to Case et al.; U.S. Patent No. 6,824,978 to Cox III et al.; U.S. Patent No. 6,794,136 to Eisenberg et al.; U.S. Patent No. 6,785,613 to Eisenberg et al.; U.S. Patent No. 6,777,185 to Case et al.; U.S. Patent No. 6,706,470 to Choo et al.; U.S. Patent No. 6,607,882 to Cox IM et al.; U.S. Patent No. 6,599,692 to Case et al.; U.S. Patent No. 6,534,261 to Cox II et al.; U.S. Patent No. 6,503,717 to Case et al.; U.S. Patent No. 6,453,242 to Eisenberg et al.; United States Patent Application Publication No. 2006/0246588 to Rebar et al.; United States Patent Application Publication No. 2006/0246567 to Rebar et al.; United States Patent Application Publication No. 2006/0166263 to Case et al.; United States Patent Application Publication No. 2006/0078878 to Cox HI et al.; United States Patent Application Publication No. 2005/0257062 to Rebar et al.; United States Patent Application Publication No. 2005/0215502 to Cox III et al.; United States Patent Application Publication No. 2005/0130304 to Cox MI et al.; United States Patent Application Publication No. 2004/0203064 to Case et al.; United States Patent Application Publication No. 2003/0166141 to Case et al.; United States Patent Application Publication No. 2003/0134318 to Case et al.; United States Patent Application Publication No. 2003/0105593 to Eisenberg et al.; United States Patent Application Publication No. 2003/0087817 to Cox IM et al.; United States Patent Application Publication No. 2003/0021776 to Rebar et al.; and United States Patent Application Publication No. 2002/0081614 to Case et al.. For example, one alternative described in these patents and patent publications involves the use of so-called "D-able sites" and zinc finger modules or zinc finger DNA binding domains that can bind to such sites. A "D-able" site is a region of a target site that allows an appropriately designed zinc finger module or zinc finger DNA binding domain to bind to four bases rather than three of the target strand. Such a zinc finger module or zinc finger DNA binding domain binds to a triplet of three bases on one strand of a double-stranded DNA target segment (target strand) and a fourth base on the other, complementary, strand. Binding of a single zinc finger to a four base target segment imposes constraints both on the sequence of the target strand and on the amino acid sequence of the zinc finger.

[0048] As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with

the standard single-letter designations used routinely in the art.

[0049] In a peptide or protein, suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, e.g. Watson et al. *Molecular Biology of the Gene*, 4th Edition, 1987, Benjamin/Cummings, p. 224). In particular, such a conservative variant has a modified amino acid sequence, such that the change(s) do not substantially alter the protein's (the conservative variant's) structure and/or activity, e.g., antibody activity, enzymatic activity, or receptor activity. These include conservatively modified variations of an amino acid sequence, i.e., amino acid substitutions, additions or deletions of those residues that are not critical for protein activity, or substitution of amino acids with residues having similar properties (e.g., acidic, basic, positively or negatively charged, polar or non-polar, etc.) such that the substitutions of even critical amino acids does not substantially alter structure and/or activity. Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one exemplary guideline to select conservative substitutions includes (original residue followed by exemplary substitution): Ala/Gly or Ser; Arg/Lys; Asn/Gln or His; Asp/Glu; Cys/Ser; Gln/Asn; Gly/Asp; Gly/Ala or Pro; His/Asn or Gln; Ile/Leu or Val; Leu/Ile or Val; Lys/Arg or Gln or Glu; Met/Leu or Tyr or Ile; Phe/Met or Leu or Tyr; Ser/Thr; Thr/Ser; Trp/Tyr; Tyr/Trp or Phe; Val/Ile or Leu. An alternative exemplary guideline uses the following six groups, each containing amino acids that are conservative substitutions for one another: (1) alanine (A or Ala), serine (S or Ser), threonine (T or Thr); (2) aspartic acid (D or Asp), glutamic acid (E or Glu); (3) asparagine (N or Asn), glutamine (Q or Gln); (4) arginine (R or Arg), lysine (K or Lys); (5) isoleucine (I or Ile), leucine (L or Leu), methionine (M or Met), valine (V or Val); and (6) phenylalanine (F or Phe), tyrosine (Y or Tyr), tryptophan (W or Trp); (see also, e.g., Creighton (1984) *Proteins*, W. H. Freeman and Company; Schulz and Schirmer (1979) *Principles of Protein Structure*, Springer-Verlag). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered "conservatively modified variations" when the three-dimensional structure and the function of the protein to be delivered are conserved by such a variation.

[0050] As used herein, the term "expression vector" refers to a plasmid, virus, phagemid, or other vehicle known in the art that has been manipulated by insertion or incorporation of heterologous DNA, such as nucleic acid encoding the fusion proteins herein or expression cassettes provided herein. Such expression vectors typically contain a promoter sequence for efficient transcription of the inserted nucleic acid in a cell. The expression vector typically contains an origin of replication, a promoter, as well as specific genes that permit phenotypic selection of transformed cells.

[0051] As used herein, the term "host cells" refers to cells in which a vector can be propagated and its DNA expressed. The term also includes any progeny of the subject host cell. It is

understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. Such progeny are included when the term "host cell" is used. Methods of stable transfer where the foreign DNA is continuously maintained in the host are known in the art.

[0052] As used herein, genetic therapy involves the transfer of heterologous DNA to the certain cells, target cells, of a mammal, particularly a human, with a disorder or conditions for which such therapy is sought. The DNA is introduced into the selected target cells in a manner such that the heterologous DNA is expressed and a therapeutic product encoded thereby is produced. Alternatively, the heterologous DNA may in some manner mediate expression of DNA that encodes the therapeutic product, or it may encode a product, such as a peptide or RNA that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy may also be used to deliver nucleic acid encoding a gene product that replaces a defective gene or supplements a gene product produced by the mammal or the cell in which it is introduced. The introduced nucleic acid may encode a therapeutic compound, such as a growth factor inhibitor thereof, or a tumor necrosis factor or inhibitor thereof, such as a receptor therefor, that is not normally produced in the mammalian host or that is not produced in therapeutically effective amounts or at a therapeutically useful time. The heterologous DNA encoding the therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof. Genetic therapy may also involve delivery of an inhibitor or repressor or other modulator of gene expression.

[0053] As used herein, heterologous DNA is DNA that encodes RNA and proteins that are not normally produced in vivo by the cell in which it is expressed or that mediates or encodes mediators that alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes. Heterologous DNA may also be referred to as foreign DNA. Any DNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which is expressed is herein encompassed by heterologous DNA. Examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers drug resistance, DNA that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and DNA that encodes other types of proteins, such as antibodies. Antibodies that are encoded by heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

[0054] Hence, herein heterologous DNA or foreign DNA, includes a DNA molecule not present in the exact orientation and position as the counterpart DNA molecule found in the genome. It may also refer to a DNA molecule from another organism or species (i.e., exogenous).

[0055] As used herein, a therapeutically effective product is a product that is encoded by heterologous nucleic acid, typically DNA, that, upon introduction of the nucleic acid into a host, a product is expressed that ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease. Typically, DNA encoding a desired gene product is cloned into a plasmid vector and introduced by routine methods, such as calcium-phosphate mediated DNA uptake (see, (1981) *Somat. Cell. Mol. Genet.* 7:603-616) or microinjection, into

producer cells, such as packaging cells. After amplification in producer cells, the vectors that contain the heterologous DNA are introduced into selected target cells.

[0056] As used herein, an expression or delivery vector refers to any plasmid or virus into which a foreign or heterologous DNA may be inserted for expression in a suitable host cell-i.e., the protein or polypeptide encoded by the DNA is synthesized in the host cell's system. Vectors capable of directing the expression of DNA segments (genes) encoding one or more proteins are referred to herein as "expression vectors". Also included are vectors that allow cloning of cDNA (complementary DNA) from mRNAs produced using reverse transcriptase.

[0057] As used herein, a gene refers to a nucleic acid molecule whose nucleotide sequence encodes an RNA or polypeptide. A gene can be either RNA or DNA. Genes may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

[0058] As used herein, the term "isolated" with reference to a nucleic acid molecule or polypeptide or other biomolecule means that the nucleic acid or polypeptide has been separated from the genetic environment from which the polypeptide or nucleic acid were obtained. It may also mean that the biomolecule has been altered from the natural state. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated," as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of a compound can be substantially purified by the one-step method described in Smith et al. (1988) Gene 67:3140. The terms isolated and purified are sometimes used interchangeably.

[0059] Thus, by "isolated" is meant that the nucleic acid is free of the coding sequences of those genes that, in a naturally-occurring genome immediately flank the gene encoding the nucleic acid of interest. Isolated DNA may be single-stranded or double-stranded, and may be genomic DNA, cDNA, recombinant hybrid DNA, or synthetic DNA. It may be identical to a native DNA sequence, or may differ from such sequence by the deletion, addition, or substitution of one or more nucleotides.

[0060] "Isolated" or "purified" as those terms are used to refer to preparations made from biological cells or hosts means any cell extract containing the indicated DNA or protein including a crude extract of the DNA or protein of interest. For example, in the case of a protein, a purified preparation can be obtained following an individual technique or a series of preparative or biochemical techniques and the DNA or protein of interest can be present at various degrees of purity in these preparations. Particularly for proteins, the procedures may include for example, but are not limited to, ammonium sulfate fractionation, gel filtration, ion exchange chromatography, affinity chromatography, density gradient centrifugation, electrofocusing, chromatofocusing, and electrophoresis.

[0061] A preparation of DNA or protein that is "substantially pure" or "isolated" should be understood to mean a preparation free from naturally occurring materials with which such DNA or protein is normally associated in nature. "Essentially pure" should be understood to mean a "highly" purified preparation that contains at least 95% of the DNA or protein of interest.

[0062] A cell extract that contains the DNA or protein of interest should be understood to mean a homogenate preparation or cell-free preparation obtained from cells that express the protein or contain the DNA of interest. The term "cell extract" is intended to include culture media, especially spent culture media from which the cells have been removed.

[0063] As used herein, a promoter region of a gene includes the regulatory element or elements that typically lie 5' to a structural gene; multiple regulatory elements can be present, separated by intervening nucleotide sequences. If a gene is to be activated, proteins known as transcription factors attach to the promoter region of the gene. This assembly resembles an "on switch" by enabling an enzyme to transcribe a second genetic segment from DNA into RNA. In most cases the resulting RNA molecule serves as a template for synthesis of a specific protein; sometimes RNA itself is the final product. The promoter region may be a normal cellular promoter or, for example, an onco-promoter. An onco-promoter is generally a virus-derived promoter. Viral promoters to which zinc finger binding polypeptides may be targeted include, but are not limited to, retroviral long terminal repeats (LTRs), and Lentivirus promoters, such as promoters from human T-cell lymphotropic virus (HTLV) 1 and 2 and human immunodeficiency virus (HIV) 1 or 2.

[0064] As used herein, the term "truncated" or similar terminology refers to a polypeptide derivative that contains less than the full amino acid sequence of a native protein, such as a ZFP, TALE or serine recombinase.

[0065] As used herein, a polypeptide "variant" or "derivative" refers to a polypeptide that is a mutagenized form of a polypeptide or one produced through recombination but that still retains a desired activity, such as the ability to bind to a ligand or a nucleic acid molecule or to modulate transcription.

[0066] As used herein, the terms "pharmaceutically acceptable", "physiologically tolerable" and grammatical variations thereof, as they refer to compositions, carriers, diluents and reagents, are used interchangeably and represent that the materials are capable of administration to or upon a human without the production of undesirable physiological effects such as nausea, dizziness, gastric upset and the like which would be to a degree that would prohibit administration of the composition.

[0067] As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. Preferred vectors are those capable of autonomous replication and expression of structural gene products present in the DNA segments to which they are operatively linked. Vectors, therefore, preferably contain the replicons and selectable markers described earlier. Vectors include, but are not necessarily limited to, expression vectors.

[0068] As used herein with regard to nucleic acid molecules, including DNA fragments, the phrase "operatively linked" means the sequences or segments have been covalently joined, preferably by conventional phosphodiester bonds, into one strand of DNA, whether in single or double-stranded form such that operatively linked portions function as intended. The choice of vector to which transcription unit or a cassette provided herein is operatively linked depends directly, as is well known in the art, on the functional properties desired, e.g., vector replication and protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules.

[0069] As used herein, administration of a therapeutic composition can be effected by any means, and includes, but is not limited to, oral, subcutaneous, intravenous, intramuscular, intrasternal, infusion techniques, intraperitoneal administration and parenteral administration.

[0070] Methods of transforming cells are well known in the art. By "transformed" it is meant a heritable alteration in a cell resulting from the uptake of foreign DNA. Suitable methods include viral infection, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. in vitro, ex vivo, or in vivo). A general discussion of these methods can be found in Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

[0071] The terms "nucleic acid molecule" and "polynucleotide" are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. Non-limiting examples of polynucleotides include a gene, a gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers.

[0072] An "expression cassette" comprises any nucleic acid construct capable of directing the expression of a gene/coding sequence of interest. Such cassettes can be constructed into a "vector," "vector construct," "expression vector," or "gene transfer vector," in order to transfer the expression cassette into target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

[0073] Techniques for determining nucleic acid and amino acid "sequence identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of

two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, Wis.). A preferred method of establishing percent identity in the context of the present invention is to use the MPSRCE package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, Calif.). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR.

[0074] Alternatively, homology can be determined by hybridization of polynucleotides under conditions that form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 80%-85%, preferably at least about 85%-90%, more preferably at least about 90%-95%, and most preferably at least about 95%-98% sequence identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; DNA Cloning, *supra*; Nucleic Acid Hybridization, *supra*.

[0075] As such, the invention provides nucleic acid and amino acid sequences encoding chimeric polypeptides of the invention which are substantially homologous and encode polypeptides that retain equivalent biological activity.

[0076] Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and

strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

[0077] When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, *Nucleic Acid Hybridization: A Practical Approach*, editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, D.C.; IRL Press).

[0078] With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.)

[0079] A first polynucleotide is "derived from" a second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

[0080] A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

[0081] Site-specific recombinases are powerful tools for genome engineering. Hyperactivated variants of the resolvase/invertase family of serine recombinases function without accessory factors, and thus can be re-targeted to sequences of interest by replacing native DNA-binding domains with engineered zinc-finger proteins (ZFPs).

[0082] The zinc finger recombinases described herein are chimeric enzymes composed of an activated catalytic domain derived from the resolvase/invertase family of serine recombinases and a custom- designed zinc-finger DNA-binding domain. The ZFRs assembled from engineered catalytic domains efficiently recombine user-defined DNA targets with high specificity and designed ZFRs integrate DNA into targeted endogenous loci in human cells.

[0083] In one aspect, the invention provides a method of generating a plurality of zinc finger recombinase (ZFRs) proteins having catalytic specificity greater than the corresponding wild type recombinase. The method includes performing random mutagenesis on a recombinase catalytic domain at positions equivalent to Gin Ile120, Thr123, Leu127, Ile136 and Gly137 or a combination thereof with reference to a wild-type Gin catalytic domain, mutating the DNA at positions 2 and 3 for each amino acid; fusing the recombinase catalytic domain with a plurality of zinc finger binding domains to form ZFRs, and enriching for ZFRs having catalytic specificity greater than the corresponding wild type recombinase. In embodiments the ZFRs have increased catalytic activity on DNA targets selected from GC, GT, CA, TT and AC. In one embodiment, the recombinase catalytic domain is mutagenized at Ile136 and/or Gly137.

[0084] As used herein, a wild-type Gin catalytic domain refers to a Gin catalytic domain including all or a portion of a polypeptide having the amino acid sequence set forth as SEQ ID NO: 56 as follows:
MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD
TLVVKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAE
MERELIERTMAGLAAARNKGRIGGRPPKLTKAWEQAGRLLAQGIPRKQVALIYDV
ALSTLYKKHP

[0085] In variousembodiments, the chimeric polypeptides of the invention include a Gin catalytic domain, such as those generated by the method of the invention. Particular Gin catalytic domains include those set forth in Table 1.

Table 1. Gin catalytic domains.

Gin catalytic domains.		
Variant	SEQ ID NO:	Sequence
Gin α	57	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAE MERELIERTMAGLAAARNKGRIGGRPPKSG
Gin β	58	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS

Gin catalytic domains.		
Variant	SEQ ID NO:	Sequence
		SPMGRFFFYVMGALAEMERELIERMAGIAAARNKGRIRGRPPKS
Gin γ	59	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRITFYVMGALAEMERELIERMAGIAAARNKGRIRGRPPKSG
Gin δ	60	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIERMAGIAAARNKGRIRGRPPKSG
Gin ε	61	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIERMAGIAAARNKGRIRGRPPKSG
Gin ζ	62	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRA LKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTS SPMGRFFFYVMGALAEMERELIERTMAGIAAARNKGRIRGRPPKSG
Targeted arm region positions are highlighted. Random substitutions are emboldened and underlined. The hyperactivating H106Y mutation is underlined.		

[0086] In various embodiments, the ZFRs generated by the method of the invention include a Gin catalytic domain operatively linked to a plurality of zinc finger binding domains. Exemplary ZFRs generated by the invention include those set forth in Table 2.

Table 2. ZFRs.

Amino acid sequences of exemplary ZFRs.		
ZFR-1 Left	SEQ ID NO: 63	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRITFYVMGALAEMERELIERMAGIAAARNKGRIRGR RPPKSGTGEKPYKCPECGKSFSTSGNLVRHQRTHTGEKPYKCPECG KSFSQSGDLRRHQRTHTGEKPYKCPECGKSFSTSGNLVRHQRTHTG EKPYKCPECGKSFSTSGELVRHQRTHTGKKTSGQAGQ
ZFR-1 Right	SEQ ID NO: 64	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTD

Amino acid sequences of exemplary ZFRs.		
		SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFS <u>SHRTTLTNH</u> QRTHTGEKPYKCPECG KSFS <u>QSGDLRRH</u> QRTHTGEKPYKCPECGKSFS <u>QSGDLRRH</u> QRTHTG EKPYKCPECGKSFS <u>QSGDLRRH</u> QRTHTGKKTSQGAGQ
ZFR-2 Left	SEQ ID NO: 65	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIIFDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFS <u>QSGDLRRH</u> IQRTIITGEKPYKCPECG KSFS <u>QRAHLERH</u> QRTHTGEKPYKCPECGKSFS <u>STSGNLVRH</u> QRTHTG EKPYKCPECGKSFS <u>RSDELVRH</u> QRTHTGKKTSQGAGQ
ZFR-2 Right	SEQ ID NO: 66	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIIFDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFS <u>RSDKLVRH</u> QRTHTGEKPYKCPECG KSFS <u>RKDNLKNH</u> QRTHTGEKPYKCPECGKSFS <u>STSGELVRH</u> QRTHTG EKPYKCPECGKSFS <u>RSDKLVRH</u> QRTHTGKKTSQGAGQ
ZFR-3 Left	SEQ ID NO: 67	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIIFDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFS <u>TTGNLTVH</u> QRTHTGEKPYKCPECG KSFS <u>DPGALVRH</u> QRTHTGEKPYKCPECGKSFS <u>QSSNLVRH</u> QRTHTG EKPYKCPECGKSFS <u>RSDHLTNH</u> QRTHTGKKTSQGAGQ
ZFR-3 Right	SEQ ID NO: 68	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIIFDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG RPPKSGTGEKPYKCPECGKSFS <u>RKDNLKNH</u> QRTHTGEKPYKCPECG KSFS <u>RSDIILTNI</u> QRTIITGEKPYKCPECGKSFS <u>DPGNLVRH</u> IQRTIITG EKPYKCPECGKSFS <u>RKDNLKNH</u> QRTHTGKKTSQGAGQ
ZFR-4 Left	SEQ ID NO: 69	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIIFDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDRLGSRMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELIERIMAGIAAARNKGRIGG GRPPKSGTGEKPYKCPECGKSFS <u>QANLRAH</u> QRTHTGEKPYKCPECG GKSFS <u>QSSSLVRH</u> QRTHTGEKPYKCPECGKSFS <u>TTGNLTVH</u> QRTHT

Amino acid sequences of exemplary ZFRs.		
		GEKPYKCPECGKSFSQRAII ~~ER~~ IIQRTIITGKKTSGQAGQ
ZFR-4 Right	SEQ ID NO: 70	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDR ~~LGR~~ SMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELI ~~ER~~ ~~V~~ MAG ~~I~~ AAARNKGR ~~W~~ GRPPKSGTGEKPYKCPECGKSFSQ ~~RA~~ NLRAIIQRTIITGEKPYKCPEC GKSFSR ~~DEL~~ NVHQRTHTGEKPYKCPECGKSFSQ ~~LA~~ HLRAHQRTHT GEKPYKCPECGKSFSQ ~~RA~~ HLERHQRTHTGKKTSGQAGQ
ZFR-5 Left	SEQ ID NO: 71	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDR ~~LGR~~ SMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELI ~~ER~~ ~~V~~ MAG ~~I~~ AAARNKGR ~~W~~ GRPPKSGTGEKPYKCPECGKSFSR ~~DEL~~ NVHQRTHTGEKPYKCPEC GKS ~~I~~ S ~~R~~ S ~~D~~ H ~~L~~ T ~~N~~ HQRTHTGEKPYKCPLCGKSI ~~'~~ SQ ~~LA~~ HLRAHQRTHT GEKPYKCPECGKSFSQ ~~RA~~ HLERHQRTHTGKKTSGQAGQ
ZFR-5 Right	SEQ ID NO: 72	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDR ~~LGR~~ SMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELI ~~ER~~ ~~V~~ MAG ~~I~~ AAARNKGR ~~I~~ G RPPKSGTGEKPYKCPECGKSFS ~~T~~ SGSLVRHQRTHTGEKPYKCPEC KSI ~~'~~ S ~~R~~ S ~~D~~ K ~~L~~ VRHQRTHTGEKPYKCPECGKSFSQ ~~S~~ GDLRRHQRTHTG EKPYKCPECGKSFS ~~T~~ SGELVRHQRTHTGKKTSGQAGQ
ZFR-6 Left	SEQ ID NO: 73	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDR ~~LGR~~ SMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELI ~~ERT~~ SAG ~~I~~ AAA ~~I~~ NKGRIMGR ~~P~~ KSGTGEKPYKCPECGKSFSQ ~~LA~~ HLRAHQRTHTGEKPYKCPECGK SFSQ ~~LA~~ AILRAIIQRTIITGEKPYKCPECGKSFS ~~D~~ PGIILVRIHQRTIITGE KPYKCPECGKSFS ~~D~~ SGNLRVHQRTHTGKKTSGQAGQ
ZFR-6 Right	SEQ ID NO: 74	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWKLDR ~~LGR~~ SMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFYVMGALAEMERELI ~~ER~~ ~~V~~ MAG ~~I~~ AAARNKGR ~~W~~ GRPPKSGTGEKPYKCPECGKSFSQ ~~RA~~ HLERHQRTHTGEKPYKCPEC GKSFS ~~T~~ TGNLTVIIHQRTIITGEKPYKCPECGKSFS ~~D~~ SGNLRV ~~I~~ IQRTIIT GLKPYKCPLCGKSI ~~'~~ SQ ~~S~~ SNLVRHQRTHTGKKTSGQAGQ
ZFR-7	SEQ ID	

Amino acid sequences of exemplary ZFRs.		
Left	NO: 75	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK
		RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERTMAGIAAARNKGRIRGR PPKSGTGEKPYKCPECGKSFSTHLDLIRHQRTHTGEKPYKCPECGKS FSTTGNLTVHQRTHTGEKPYKCPECGKSFSSQSSSLVRHQRTHTGEKP YKCPECGKSFSDNLVRHQRTHTGKKTSQGAGQ
ZFR-7 Right	SEQ ID NO: 76	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERTMAGIAAARNKGRIRGR PPKSGTGEKPYKCPECGKSFSDKILVRHQRTHTGEKPYKCPECGK SFSRRDELNVHQRTHTGEKPYKCPECGKSFSSQSSSLVRHQRTHTGE KPYKCPECGKSFSDHLTNHQRTHTGKKTSQGAGQ
ZFR-8 Left	SEQ ID NO: 77	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERIMAGIAAARNKGRIRGG RPPKSGTGEKPYKCPECGKSFSSQRAHLERHQRTHTGEKPYKCPECG KSFSSTGNLVRHQRTHTGEKPYKCPECGKSFSDDELVRHQRTHTG EKPYKCPECGKSFSSHKNALQNHQRTHTGKKTSQGAGQ
ZFR-8 Right	SEQ ID NO: 78	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERIMAGIAAARNKGRIRGG RPPKSGTGEKPYKCPECGKSFSDDELNVHQRTHTGEKPYKCPECG KSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSSQSSSLVRHQRTHTGE KPYKCPECGKSFSTTGNLTVHQRTHTGKKTSQGAGQ
ZFR-9 Left	SEQ ID NO: 79	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERIMAGIAAARNKGRIRGV GRPPKSGTGEKPYKCPECGKSFSTTGNLTVHQRTHTGEKPYKCPEC GKSFSQSSNLVRHQRTHTGEKPYKCPECGKSFSSQRAHLERHQRTHT GEKPYKCPECGKSFSSQSSLIHQRTHTGKKTSQGAGQ
ZFR-9 Right	SEQ ID NO: 80	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLK RALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTD SIDTSSPMGRFFFFYVMGALAEMERELIERTMAGIAAARNKGRIRGR PPKSGTGEKPYKCPECGKSFSDPGALVRHQRTHTGEKPYKCPECGK

Amino acid sequences of exemplary ZFRs.		
		SFSQSSSLVRHQRTHTGEKPYKCPECGKSFSQLAHLRAHQRTHTGE KPYKCPECGKSFSQ ~~QRANL~~ RAHQRTHTGKKTSQGAGQ
Arm region mutations are highlighted. Specificity-determining α -helical zinc-finger residues are underlined.		

[0087] While the Examples illustrate generation of ZFRs having a Gin catalytic domain, the methods may be applied to catalytic domains of a number of other recombinases. Such recombinases include: a) Tn3, also known as EcoTn3; Hin, also known as StyHin; MuGin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsoISC1904a; SsoISC1904b; SsoISC1913; Aam606; MjaM0014; Pab; HpyIS607; MtuIS_Y349; MtuRv2792c; MtuRv2979c; MtuRv3828c; MtuRv0921; MceRv0921; TnpX; TndX; WwK; lactococcal phage TP901-1 serine recombinase; *S. pyogenes* phage ϕ 370.1 serine recombinase; *S. pyogenes* phage ϕ FC1 serine recombinase; *Listeria* phage A118 serine recombinase; *S. coelicolor* chromosome SC3C8.24 serine recombinase; *S. coelicolor* chromosome SC2E1.37 serine recombinase; *S. coelicolor* chromosome SCD78.04c serine recombinase; *S. coelicolor* chromosome SC8F4.15c serine recombinase; *S. coelicolor* chromosome SCD12A.23 serine recombinase; *S. coelicolor* chromosome SCH10.38c serine recombinase; *S. coelicolor* chromosome SCC88.14 serine recombinase; *Streptomyces* phage ϕ C31 serine recombinase; *Streptomyces* phage R4 serine recombinase; *Bacillus* phage ϕ 105 serine recombinase; *Bacillus* phage SPBc2 serine recombinase; *Bacillus* prophage SKIN serine recombinase; *S. aureus ccrA* serine recombinase; *S. aureus ccrB* serine recombinase; *M. tuberculosis* phage Bxb1 serine recombinase; *M. tuberculosis* prophage ϕ RV1 serine recombinase; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; and Cac 1954; and b) muteins of a).

[0088] Imperfect modularity with particular domains, lack of high-affinity binding to all DNA triplets, and difficulty in construction has hindered the widespread adoption of ZFPs in unspecialized laboratories. The discovery of a novel type of DNA-binding domain in transcription activator-like effector (TALE) proteins from *Xanthomonas* provides an alternative to ZFPs. Described herein are chimeric TALE recombinases (TALERS): engineered fusions between a hyperactivated catalytic domain from the DNA invertase Gin and an optimized TALE architecture. A library of incrementally truncated TALE variants was identified to identify TALER fusions that modify DNA with efficiency and specificity comparable to zinc-finger recombinases in bacterial cells. Also shown in the Examples, TALERS recombine DNA in mammalian cells. The TALER architecture described herein provides a platform for insertion of customized TALE domains, thus significantly expanding the targeting capacity of engineered recombinases and their potential applications in biotechnology and medicine.

[0089] Transcription activator-like effector (TALE) proteins can be designed to bind virtually any DNA sequence. General guidelines for design of TALE DNA-binding domains suggest that the 5'-

most base of the DNA sequence bound by the TALE (the N_0 base) should be a thymine. We quantified the N_0 requirement by analysis of the activities of TALE transcription factors (TALE-TF), TALE recombinases (TALE-R) and TALE nucleases (TALENs) with each DNA base at this position. In the absence of a 5' T, we observed decreases in TALE activity up to >1000-fold in TALE-TF activity, up to 100-fold in TALE-R activity and up to 10-fold reduction in TALEN activity compared with target sequences containing a 5' T. To develop TALE architectures that recognize all possible N_0 bases, a structure-guided library design coupled with TALE-R activity selections was used to evolve novel TALE N-terminal domains to accommodate any N_0 base. A G-selective domain and broadly reactive domains were isolated and characterized. The engineered TALE domains selected in the TALE-R format demonstrated modularity and were active in TALE-TF and TALEN architectures. Evolved N-terminal domains provide effective and unconstrained TALE-based targeting of any DNA sequence as TALE binding proteins and designer enzymes.

[0090] In one aspect, the invention provides a method of generating a transcription activator-like effector (TALE) protein binding domain which specifically binds a desired nucleotide. As shown in the Examples, the method includes a) randomizing the amino acid sequence of the TALE protein binding domain by mutating an amino acid residue within a variable di-residue (RVD), or within 1 to 2 amino acid residues N-terminal or C-terminal of the RVD; and b) selecting for the randomized TALE protein binding domain of (a), wherein the TALE protein binding domain specifically binds to the desired nucleotide.

[0091] Sequence-specific nucleases, recombinases, nucleases and transcription factors are provided herein. The sequence-specific polypeptides include customized TAL effector DNA binding domains. As such, in another aspect, the invention provides a chimeric polypeptide. The polypeptide includes: a) a recombinase, a transcription factor or nuclease; and b) a transcription activator-like effector (TALE) protein.

[0092] TALEs are proteins of plant pathogenic bacteria that are injected by the pathogen into the plant cell, where they travel to the nucleus and function as transcription factors to turn on specific plant genes. The primary amino acid sequence of a TALE dictates the nucleotide sequence to which it binds. Thus, target sites can be predicted for TALE, and TALE also can be engineered and generated for the purpose of binding to particular nucleotide sequences, as described herein.

[0093] Fused to the TALE-encoding nucleic acid sequences are sequences encoding a nuclease, transcription factor or recombinase, or a portion thereof. Many such proteins are known in art that may be used in the present invention.

[0094] In various embodiments, the chimeric polypeptide includes a catalytic domain of a recombinase. As discussed above, catalytic domains of a number of recombinases may be utilized. Such recombinases include: a) Tn3, also known as EcoTn3; Hin, also known as StyHin; Gin, also known as MuGin; Sin; Beta; Pin; Min; Din; Cin; EcoTn21; SfaTn917; BmeTn5083; Bme53; Cpe; SauSK1; SauSK41; SauTn552; Ran; Aac; Lla; pMER05; Mlo92; Mlo90; Rrh; Pje; Req; PpsTn5501; Pae; Xan; ISXc5; Spy; RhizY4cG; SarpNL1; SsoISC1904a; SsoISC1904b; SsoISC1913; Aam606; MjaM0014; Pab; HpyIS607; MtuIS_Y349; MtuRv2792c; MtuRv2979c;

MtuRv3828c; MtuRv0921; MceRv0921; TnpX; TndX; WwK; lactococcal phage TP901-1 serine recombinase; *S. pyogenes* phage ϕ 370.1 serine recombinase; *S. pyogenes* phage ϕ FC1 serine recombinase; *Listeria* phage A118 serine recombinase; *S. coelicolor* chromosome SC3C8.24 serine recombinase; *S. coelicolor* chromosome SC2E1.37 serine recombinase; *S. coelicolor* chromosome SCD78.04c serine recombinase; *S. coelicolor* chromosome SC8F4.15c serine recombinase; *S. coelicolor* chromosome SCD12A.23 serine recombinase; *S. coelicolor* chromosome SCH10.38c serine recombinase; *S. coelicolor* chromosome SCC88.14 serine recombinase; *Streptomyces* phage ϕ C31 serine recombinase; *Streptomyces* phage R4 serine recombinase; *Bacillus* phage ϕ 105 serine recombinase; *Bacillus* phage SPBc2 serine recombinase; *Bacillus* prophage SKIN serine recombinase; *S. aureus* *ccrA* serine recombinase; *S. aureus* *ccrB* serine recombinase; *M. tuberculosis* phage Bxb1 serine recombinase; *M. tuberculosis* prophage ϕ RV1 serine recombinase; YBCK_ECOLI; Y4bA; Bja; Spn; Cac 1956; and Cac 1954; and b) muteins of a). In preferred embodiments, a highly active Gin catalytic domain is utilized. Such a domain may be generated using the methods of the present invention as described herein.

[0095] As described herein, TALEs may include a number of imperfect repeats that determine the specificity with which they interact with DNA. Each repeat binds to a single base, depending on the particular di-amino acid sequence at residues 12 and 13 of the repeat. Thus, by engineering the repeats within a TALE, particular DNA sites can be targeted. Such engineered TALEs can be used, for example, as transcription factors targeted to particular DNA sequences.

[0096] As illustrated in the Examples, the chimeric proteins of the present invention are exemplified by the variants and portions thereof (e.g., RVDs and NTDs) as set forth in Table 3.

Table 3.

Variant	SEQ ID NO:	Sequence
TALEN (Goldy) NT-T1 Protein Sequence	81	MRSPKKKRKVQVDLRTLGYSSQQQEKIKPKVRSTVAQHH EALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEAT HEDIVGVGKQWSGARALEALLTDAGELRGPPLQLDTGQLV KIAKRGGVTAMEAVIIASRNALTGAPLNLTDPQVVAIASNG GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQR LLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLC QDIIGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDIIGL PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALET VQRLLPVLCQDIIGLTPDQVVAIVSHDGGKQALETVQRLLP VLCQDHGLTPDQVVAIVSNGGKQALETVQRLLPVLCQD HGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPD QVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIA SNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGK QALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALES IVAQLSRPDPALAALTNDHLVALACLGGRPAMDVKKGLP HAPELIRRVNRRIGERTSHRVAGSQLVKSELEKKSELRHK LKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGK HLGGSRKPDGAIYTVGSPIDYGVIIVDTKAYSGGYNLPIGQA DEMORYVEENQTRNKHINPNEWVKVYPSSVTEFKFLFVSG HFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGT LTLEEVRRKFNNGEINF = N-Terminal Domain (NTD) -varied as shown below
TALEN RVD Sequences		
G1	82	NG-NN-HD-NG-HD-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-HD- NI targeting (TCTTCATTACACCTGCA; SEQ ID NO: 280)
G2	83	HD-NI-NN-NG-HD-NI-NN-NG-NI-NG-HD-NI-NI-NG-NG targeting (CAGTCAGTATCAATT; SEQ ID NO: 281)
A1	84	HD-HD-NG-NN-HD-NI-NN-HD-NG-HD-NG-HD-NI-NG-NG-NG- NG targeting (CCTGCAGCTCTTCATTCTT; SEQ ID NO: 282)
A2	85	NI-NG-NG-HD-NG-NG-HD-HD-NI-NN-NI-NG-NG-NN-NI targeting (ATTCTTCCAGAATTGA; SEQ ID NO: 283)

TALEN RVD Sequences		
C2	86	HD-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG-NN-NI-HD-NG targeting (CAGAATTGATACTGACT; SEQ ID NO: 284)
T1	87	NG-HD-NI-NG-NG-NI-HD-NI-HD-HD-NG-NN-HD-NI-NN-HD targeting (TCATTACACCTGCAGC; SEQ ID NO: 285)
T2	88	HD-NG-NG-HD-HD-NI-NN-NI-NI-NG-NG-NN-NI-NG-NI-HD-NG- NN targeting (CTTCCAGAATTGATACTG; SEQ ID NO: 286)
N-Terminal Domains		
NTD = dHax3-TALEN DNA Sequence	89	ATGAGATCTCCTAAGAAAAAGAGGAAGATGGTGGACTTGA GGACACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAA GCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCG CTTGTGGGGCATGGCTTCACTCATGCGCATATGTTCGCGCTT TCACAGCACCCCTGCGGCGCTTGGGACGGTGGCTGTCAAATA CCAAGATATGATTGCGGCCCTGCCCCAAGCCACGCACGAGG CAATTGTAGGGGTGCGTAAACAGTGGTCGGGAGCGCGAGC ACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGGGC CTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCG AAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC
NTD = dHax3-TALEN	90	MRSPKKKRKMVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGHGFTTHAHIVALSQHPAALGTVAVKYQDMIAALPLEATHEAI
		VGVGKQWSGARALEALITVAGELRGPPIQLDTGQILKIAKRG GVTAVEAVHASRNALTGAP... repeat variable diresidues
NTD = NT-βN TALEN DNA Sequence	91	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCCACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATATCATGGGGCACGCGC TCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTC CGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGATTGCA

N-Terminal Domains		
		AAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCATGCAT CGCGCAATGCACTGACGGGTGCCCCC
NTD = NT-βN TALEN	92	MRSPKKKRKVQVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGIIGFTIIAHIVALSQIIPAALGTVAVTYQIIITALPEATHIEDIV GVGKYHGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGV TAMEAVHASRNALTGAP.... repeat variable diresidues
NTD NT-G TALEN DNA Sequence	93	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAATCGCGGTCGGGGGCAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD NT-G TALEN Protein Sequence	94	MRSPKKKRKVQVDLRTLGYSSQQQEKIKPKVRSTVAQHHEAL VGIIGFTIIAHIVALSQIIPAALGTVAVTYQIIITALPEATHIEDIV GVGKSRSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGG VTAMEAVHASRNALTGAP... repeat variable diresidues
NTD = NT-αN TALEN DNA Sequence	95	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACCACGGG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACGGGGGGCTGGTGCAC GCGCTCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
NTD = NT-αN TALEN Protein	96	MRSPKKKRKVQVDLRTLGYSSQQQEKIKPKVRSTVAQIIIGA LVGHGFTHAHIVALSQHPAALGTVAVTYQHIIITALPEATHEDI repeat variable diresidues

N-Terminal Domains		
		VGVGKRGAGARALEALLTDAGETLRGPPTQLDTGQLVRIAKRG GVTAMEAVHASRNALTGAP...repeat variable diresidues
NTD = NT-T T-1 TALEN DNA	97	ATGAGATCTCCTAAGAAAAAGAGGAAGGTGCAGGTGGATC TACGCACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGAT CAAACCGAAGGTGCGTTCGACAGTGGCGCAGCACACGAG GCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCA CGTATCAGCACATAATCACGGCGTTGCCAGAGGCGACACAC GAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCAC GCGCCCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGA GGTCCGCCGTTACAGTTGGACACAGGCCAACTTGTGAAGAT TGCAAAACGTGGCGGCGTGACCGCAATGGAGGCAGTGCAT GCATCGCGCAATGCACTGACGGGTGCCCCC
MBP-TALE Protein Sequence	98	MKIEEGKLVIWINGDKGYNGLAEVGGKFEKDTGIKVTVEHPD KLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLIAETPDKA IQDKLYPIFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPK TWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIADGGYAF KYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDY SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKG QPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNK DKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIQMS AFWYAVRTAVINAASGRQTVDEALKDAQTNSSNNNNNNNN NNLGIEGRISEFGSPARPPRAKPAPRRRSAQPSDASPAQVDLR
		TLGYSQQQKEKIKPKVRSTVAQHIIIEALVGIIGFTIIHIVALSQ HPAALGTVAVTYQHIITALPEATHEDIVGVGK[XXX]GARALE ALLTDAGELLRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRN ALTGAPLNLTDPQVVAIASNIGGKQALETVQRLLPVLCQDIIGL TPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA LETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLP VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL TPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRL LPVLCQDHGLTPDQVVAIVSNGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQ LSRPDPALAALTNDHLVALACLG

N-Terminal Domains		
		XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
TALE-R Protein Sequence	99	MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDR PGLKRAIKRIQKGDTLVVWKLDRI.GRSMKHLISLVGEI.RERG INFRSLTDSIDTSSPMGRITTYVMGALAEMERELIERTMAGLA AARNKGRIGGRPPKSGSPRPPRAKPAPRRRAAQPSDASPAAQV DLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVA LSQIIPAALGTVAVTYQIIITALPEATHIEDIVGVGK[XXX]GARA LEALLTDAGELRGPPQLDITGQLVKIAKRGGVTAMEAVHASR NALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVA IASNIGGKQALETVQRLLPVLCQDIIGLTPDQVVAIASNIGGKQ ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLL PVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVA IASHDGGKQALETVQRLLPVLCQDIIGLTPDQVVAIVSIHDGGK QALETVQRLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQR LLPVLCQDHGLTPDQVVAIVSNNGGGKQALETVQRLLPVLCQD HGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQV VAIASHDGGKQALETVQRLLPVLCQDIIGLTPDQVVAIASNIGG KQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVA
		QLSRPDPAALAAITNDHLVALACLG XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
Avr15 TALE-TF Protein Sequence	100	MAQAASGSPRPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSS QQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAAL GTVAVTYQHIIITALPEATHIEDIVGVGK[XXX]GARALEALLTDA GELRGPPQLDITGQLVKIAKRGGVTAMEAVHASRNALTGAPL NLTPDQVVAIASNIGGKQALETVQRLLPVLCQDIIGLTPDQVV AIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK QALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRL LPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDH GLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVV AIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGG KQALETVQRLLPVLCQDIIGLTPDQVVAIVSIHDGGKQALETVQ RLLPVLCQDHGLTPDQVVAIVSHDGGKQALETVQRLLPVLCQ DHGLTPDQVVAIVSNNGGGKQALETVQRLLPVLCQDHGLTPDQ VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHD GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALET VQRLLPVLCQDHGLTPDQVVAIASNIGGKQALESIVAQLSRPD

N-Terminal Domains		
		PALAAALTNDHLVALACLGGRPAMDAVKKGLPHAPELIRRVNR RIGERTSIIRVADYAQVVRVLEFFQCHSIIPAYAFDEAMTQFGM SGQAGQASPKKKRKVGRADALDDFDLDMLGSDALDDFDLDM LGSDALDDFDLDMLGSDALDDFDLDMLINYPYDVPDYAS XXX: NT-T = QWS NT-G = SRS; NT-αN = RGA; NT-βN = Y-H
Avr15 RVD Sequence (for TALE-R, TALE- TF, MBP-TALE)	101	NI-NG-NI-NI-NI-HD-HD-HD-HD-NI-HD-HD-NI-NI targeting (ATAAACCCCTCCAA; SEQ ID NO: 287)

[0097] In various embodiments, chimeric protein includes a TALE protein having a C-terminal or N-terminal truncation. For example, the TALE protein may include all or a portion of SEQ ID NO: 2. In embodiments, the TALE protein is truncated between amino acid residues 27 and 268, 92 and 134, 120 and 129, 74 and 147, or 87 and 120, such at amino acid residue 28, 74, 87, 92, 95, 120, 124, 128, 129, 147 and 150.

[0098] In another embodiment, a isolated polypeptide comprising a transcription activator-like effector (TALE) protein is provided in which the TALE protein has an N-terminal domain (NTD) comprising an amino acid sequence as set forth in SEQ ID NO: 3 (VGKQWSGARAL) having one or more mutations or deletions selected from: Q is Y, Q is S, Q is R, W is R, W is G, W is deleted, S is R, S is H, S is A, S is N, and S is T.

[0099] In some embodiments, the NTD comprises an amino acid sequence selected from: VGKYRGARAL (SEQ ID NO: 4), VGKSRSGARAL (SEQ ID NO: 5), VGKYHGARAL (SEQ ID NO: 6), and VGKRGAGARAL (SEQ ID NO: 7).

[0100] In another embodiment, an isolated polypeptide comprising a transcription activator-like effector (TALE) protein is provided in which the TALE protein has an N-terminal domain (NTD) comprising an amino acid sequence as set forth in SEQ ID NO: 8 (IVDIAR₁QR₂SGDLA) having one or more mutations or deletions selected from: R₁ is K, Q is Y, Q is S, Q is R, R₂ is W, R₂ is G, R₂ is deleted, S is R, S is H, S is A, S is N, and S is T.

[0101] In some embodiments, the NTD comprises an amino acid sequence selected from: IVDIARQWSGDLA (SEQ ID NO: 9), IVDIARYRGDLA (SEQ ID NO: 10), IVDIARSRSGDLA (SEQ ID NO: 11), IVDIARYHGDLA (SEQ ID NO: 12), and IVDIARRGAGDLA (SEQ ID NO: 13).

[0102] In another embodiment, the TALE protein includes a modified N₀ domain having an amino acid sequence set forth as follows: LTPDQLVKIAKRGGTAMEAVHASRNALTGAPLN (SEQ ID NO: 102). In various embodiments, the TALE protein includes a mutated variant in which KRGG (SEQ ID NO: 103) of SEQ ID NO: 102 is selected from LDYE (SEQ ID NO: 104), INLV (SEQ ID NO: 105), YSKK (SEQ ID NO: 106), NMAH (SEQ ID NO: 107), SPTN (SEQ ID NO: 108), SNTR (SEQ

ID NO: 109), LTTT (SEQ ID NO: 110), VADL (SEQ ID NO: 111), MVLS (SEQ ID NO: 112), YNGR (SEQ ID NO: 113), RIPR (SEQ ID NO: 114), YSKI (SEQ ID NO: 115), LTQY (SEQ ID NO: 116), YLSK (SEQ ID NO: 117), LRPN (SEQ ID NO: 118), LFTN (SEQ ID NO: 119), LLTN (SEQ ID NO: 120), EEDK (SEQ ID NO: 121), VTAM (SEQ ID NO: 122), CPSR (SEQ ID NO: 123), LTRV (SEQ ID NO: 124), KGD L (SEQ ID NO: 125), QKAL (SEQ ID NO: 126), LYLL (SEQ ID NO: 127), WISV (SEQ ID NO: 128), GDQV (SEQ ID NO: 129) and CPSR (SEQ ID NO: 130).

[0103] In another embodiment, the TALE protein includes a modified N₁ domain having an amino acid sequence set forth as follows:

MRSPKKRKVQVDLRTLGYSSQQQEKIKPKVRSTVAQHH

EALVGHGFTHAHIVALSQHPAALGTAVTYQHIITALPEATHEDIVGVGXXXXXARA

LEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTGAP (SEQ ID NO: 131). In various embodiments, XXXXX of SEQ ID NO: 131 is KRPAG (SEQ ID NO: 132) or KRPSG (SEQ ID NO: 133). Additionally, the protein may include, a E40G mutation (with reference to SEQ ID NO: 131) that exhibits enhanced activity.

[0104] In another embodiment, the TALE protein includes a repeat domain having an amino acid sequence set forth as follows: LTPDVVAISNNGGKQALETYQRLLPVLCQDGH (SEQ ID NO: 134).

In various embodiments, the TALE protein includes a mutated variant in which SNNG (SEQ ID NO: 135) of SEQ ID NO: 134 is selected from RGGG (SEQ ID NO: 136), RGGR (SEQ ID NO: 137), RGVR (SEQ ID NO: 138), KGGG (SEQ ID NO: 139), SGGG (SEQ ID NO: 140), GGRG (SEQ ID NO: 141), LGGS (SEQ ID NO: 142), MDNI (SEQ ID NO: 143), RVMA (SEQ ID NO: 144), LASV (SEQ ID NO: 145), VGTG (SEQ ID NO: 146) and QGGG (SEQ ID NO: 147).

[0105] The following examples are provided to further illustrate the advantages and features of the present invention, but are not intended to limit the scope of the invention. While they are typical of those that might be used, other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

EXAMPLE 1

CHIMERIC TALE RECOMBINASES

Experimental Summary.

[0106] This study provides the first example of a TALE recombinase (TALER). Using a library of incrementally truncated TALE domains, an optimized TALER architecture was identified that can be used to recombine DNA in bacterial and mammalian cells. Any customized TALE repeat array can be inserted into the TALER architecture described herein, thus dramatically expanding the targeting capacity of engineered recombinases for applications in biotechnology and medicine.

[0107] The following Material and Methods were utilized in this Example.

Reagents.

[0108] All enzymes were purchased from New England BioLabs unless otherwise indicated. Primer sequences are provided in Table 4.

Table 4. Primers.

Primers used in this study		
Primers for pBLA substrate construction		
AvrXa7 lac target F	SEQ ID NO: 148	TTAATTAAAGAGTCTAGAAATATAAACCCCTCC AACCAGGTGCTAACTGTAAACCATGGTTTTGGA TTAGCACCTGGTTGGAGGGGGTTTATAAGATCT AGGAGGAATTAAAAATGAG
AvrXa7 lac target R	SEQ ID NO: 149	ACTGACCTAGAGAAGCTTATATAAACCCCTCC AACCAGGTGCTAATCCAAAACCATGGTTTAC AGTTAGCACCTGGTTGGAGGGGGTTTATACTG CAGTTATTTGTACAGTTCATC
AvrXa7 N F	SEQ ID NO: 150	TTAATTAAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAAGGTTTTGGTACCAAATGTC TATAAACCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N R	SEQ ID NO: 152	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGACATTGGTACCAAACCC TTATAAACCCCTCCAACCAGGTGCTAACATGC AGTTATTTGTACAGTTCATC
AvrXa7 N RC F	SEQ ID NO: 153	TTAATTAAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATATCCAAAACCATGGTTTACAG TATAAACCCCTCCAACCAGGTGCTAAAGATC TAGGAGGAATTTAAAATGAG
AvrXa7 N RC R	SEQ ID NO: 154	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATCCAAAACCATGGTTTACA GTATAAACCCCTCCAACCAGGTGCTAACTGC AGTTATTTGTACAGTTCATC

Primers used in this study		
Primers for pBLA substrate construction		
AvrXa7 N RC +3 F	SEQ ID NO: 155	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCCAAAACCATGGTTTA CAGGGTTATAAACCCCTCCAACCAGGTGCTA AAGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC +3 R	SEQ ID NO: 277	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAACCCTGTAAACCATGGTTT TGGAAGCTATAAACCCCTCCAACCAGGTGCT AACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +6 F	SEQ ID NO: 156	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCATCCAAAACCATGGT TTACAGGGTTCCCTATAAACCCCTCCAACCAG GTGCTAAAGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC +6 R	SEQ ID NO: 157	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCCTGTAAACCATGG TTTTGGATGAAGCTATAAACCCCTCCAACCA GGTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 N RC +12 F	SEQ ID NO: 158	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAGCTTCAGCTTCATCCAAAAC CATGGTTTACAGGGTTCCGGTTCCTATAAAC CCCTCCAACCAGGTGCTAAAGATCTAGGAGGA ATTTAAAATGAG
AvrXa7 N RC +12 R	SEQ ID NO: 278	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATAGCAACCGCAACCCTGTAAA CCATGGTTTTGGATGAAGCTGAAGCTATAAAC CCCCTCCAACCAGGTGCTAACTGCAGTTATTT GTACAGTTCATC
AvrXa7 N RC -3 F	SEQ ID NO: 160	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATAAAAACCATGGTTTATATAAA CCCCCTCCAACCAGGTGCTAAAGATCTAGGAG GAATTTAAAATGAG
AvrXa7 N RC -3 R	SEQ ID	

Primers used in this study		
Primers for pBLA substrate construction		
	NO: 161	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTG GAGGGGGTTTATATAAAACCATGGTTTTTATAA ACCCCTCCAACCAGGTGCTAACTGCAGTTAT TTGTACAGTTCATC
AvrXa7 N RC GG F	SEQ ID NO: 162	TTAATTAAGAGTCTAGATTAGCACCTGGTTGG AGGGGGTTTATATCCAAAACCGGGGTTTACA GTATAAACCCCTCCAACCAGGTGCTAAAGA TCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC GG R	SEQ ID NO: 163	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATACTGTAAACCCCGGTTTT GGATATAAACCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC
AvrXa7 N 20t F	SEQ ID NO: 164	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATACGAAATATTATAAATTA TCATATAAACCCCTCCAACCAGGTGCTAA AGATCTAGGAGGAATTTAAAATGAG
AvrXa7 N RC 20t R	SEQ ID NO: 165	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATATGATAATTTATAATATT TCGTATAAACCCCTCCAACCAGGTGCTAAC TGCAGTTATTTGTACAGTTCATC
AvrXa7 32 GG F	SEQ ID NO: 166	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCGG GGTTTACAGGGTTCCTATAAACCCCTCCAA CCAGGTGCTAAAGATCTAGGAGGAATTTAA AATGAG
AvrXa7 32 GG R	SEQ ID NO: 167	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCCTGTAAACCGG GGTTTTGGATGAAGCTATAAACCCCTCCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTTC ATC
AvrXa7 32t F	SEQ ID NO: 168	TTAATTAAGAGTCTAGATTAGCACCTGGTTG GAGGGGGTTTATAGCTTCACGAAATATTATA

Primers used in this study		
Primers for pBLA substrate construction		
		AATTATCAGGTTCCCTATAAAACCCCCCTCCAAC CAGGTGCTAAAGATCTAGGAGGAATTTAAA ATGAG
AvrXa7 32t R	SEQ ID NO: 169	ACTGACCTAGAGAAGCTTTTAGCACCTGGTT GGAGGGGGTTTATAGCAACCTGATAATTTAT AATATTTCCGTGAAGCTATAAAACCCCCCTCAA CCAGGTGCTAACTGCAGTTATTTGTACAGTT CATC
Primers for pGL3Pro target site construction.		
5' pGL3 SV40 Avr.32G BglII	SEQ ID NO: 170	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCATCCAAAACCATG GTTTACAGGGTTCCTATAAAACCCCCCTCCAAC CAGGTGCTAAGCGATCTGCATCTCAATTAGT CAGC
3' pGL3 SV40 Avr.20G HindIII	SEQ ID NO: 171	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACC CTG TAA ACC ATG GTT TTG GATGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA
5' pGL3 SV40 PH4.20G6Avr BglII	SEQ ID NO: 172	TTAATTAAGAGAGATCTGCGGGAGGCGTGTC CAAAACCATGGTTTACAGGGTTCCTATAAAC CCCCTCCAACCAGGTGCTAAGCGATCTGCAT CTCAATTAGTCAGC
3' pGL3 SV40 PH4.20G6Avr HindIII	SEQ ID NO: 173	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTT ATAGCAACCCTGTA AACCATGGTTTTTGGACACGCCTCCCGCTTTG
		CAAAAGCCTAGGCCCTCCAAA
5' pGL3 SV40 Avr.44G BglII	SEQ ID NO: 174	TTAATTAAGAGAGATCTTTAGCACCTGGTTG GAGGGGGTTTATAGCTTCAGCTTCATCCAAA ACCATGGTTTACAGGGTTCGGTTCCTATAA ACCCCCCTCCAACCAGGTGCTAAGCGATCTGTC

Primers for pGL3Pro target site construction.		
		ATCTCAATTAGTCAGC
3' pGL3 SV40 Avr.44G	SEQ ID NO: 175	ACT GAC CTA GAG AAG CTT TTA GCA CCT GGT TGG AGG GGG TTTATAGCAACCGCAA CCCTG TAA ACC ATG GTT TTG GATGAAGC TGAAGCT ATA AAC CCC CTC CAA CCA GGT GCT AAT TTG CAA AAG CCT AGG CCT CCA AA
Primers for BamHI fusions		
Gin_N-term F	SEQ ID NO: 176	AGTCAGTCGAGAGCTCATGGATCCCGGCTCTA TGCTGATTGGCTATGTAAGG
Gin_N-term R	SEQ ID NO: 177	ATGCTGATATCTAGACTATCCCGATTTAGGTGG GCGACC
Gin_C-term F	SEQ ID NO: 178	AGTCAGTCGAGAGCTCATGCTGATTGGCTATGT AAGG
Gin_C-term R	SEQ ID NO: 179	TCTAGACTACGGATCCACCGATTACGCGGGGC
Primers for designed truncations		
TalR+28 Xba	SEQ ID NO: 180	ATCGCGTATCTAGACTAGCCGAGGCAGGCCAA GGCGACG
TalR+95 Xba AvrX	SEQ ID NO: 181	ATCGCGTATCTAGACTAGCTCATCTCGAACTGC GTCATG
avr n 1	SEQ ID NO: 182	GTCGCCCCGCGTAAATCGGGATCCACTGCAGAT CGGGGGGGGGC
avr n 2	SEQ ID NO: 183	GTCGCCCCGCGTAAATCGGGATCCCCCTCGCCTG CGTTCTCGGC
avr n 3	SEQ ID NO: 184	GTCGCCCCGCGTAAATCGGGATCCGATTTCGATGC CTGCCGTCCG

Primers for designed truncations		
avr n 4	SEQ ID NO: 185	GTCGCCCCGCGTAAATCGGGATCCACCGTGCGT GTCGCTGTC ~~ACTG~~
avr n 5	SEQ ID NO: 186	GTCGCCCCGCGTAAATCGGGATCCGTGGATCTAC GCACGCTCGGC
avr n 6	SEQ ID NO: 187	GTCGCCCCGCGTAAATCGGGATCCACACACGCG CACATCGTTGC
avr n 7	SEQ ID NO: 188	GTCGCCCCGCGTAAATCGGGATCCCACGAAGAC
		ATCGTTGGCGTCG
avr n 8	SEQ ID NO: 189	GTCGCCCCGCGTAAATCGGGATCCAGCGC ~~t~~ CTGG AGGCC ~~t~~ TTGC ~~TC~~
avr n 9	SEQ ID NO: 190	GTCGCCCCGCGTAAATCGGGATCCTTGGACACA GGCCAACTTCTC
avr n 10	SEQ ID NO: 191	GTCGCCCCGCGTAAATCGGGATCCAGCGGCGTG ACCGCA ~~g~~ TGGA
GinNTALPCRfusR	SEQ ID NO: 192	GGATCCCGATT ~~T~~ TACGCGGGC
Primers used for pcDNA cloning		
Nhe-SD-Gin F	SEQ ID NO: 193	ATCGTAGCAGCTAGCGCCACCATGCTGATTGGC TATGTAAG
GinGS R	SEQ ID NO: 194	GGATCCAGACCCCGATT ~~T~~ TACGCGGGC

Plasmid construction.

[0109] In order to introduce a *Bam*H1 restriction site either 5' or 3' to the Gin coding sequence, the Gin catalytic domain was PCR amplified with primers 5' Gin_N-term and 3' Gin_N-term or 5'

Gin_C-term and 3' Gin_C-term, respectively. PCR products were ligated into the *SacI* and *XbaI* restriction sites of pBluescriptII (Fermentas) to generate pB-Bam-Gin and pB-Gin-Bam. To generate the C-terminal and N-terminal TALER fusions, the AvrXa7 gene (kindly provided by Dr. B. Yang, Iowa State University) was released from pWAvrXa7 with *Bam*H1 and ligated into *Bam*H1 sites of pB-Bam-Gin and pB-Gin-Bam (41) to establish pB-Avr-Bam-Gin and pB-Gin-Bam-Avr, respectively. Correct construction of each TALER was verified by sequence analysis (Figures 6-16).

[0110] To generate N-terminal truncations of AvrXa7, AvrXa7 was PCR amplified using the Expand High Fidelity PCR System (Roche) with 5' Avr-n-(1-10) and 3' Avr +28 or 3' Avr +95 primers with the following program: 1 cycle of 3 min at 94 °C, 16 cycles of 1 min at 94 °C, 1 min at 52 °C, 6 min at 68 °C; and a final cycle of 1 hr at 68 °C. The Gin catalytic domain was PCR amplified under standard PCR conditions with 5' Gin_C-term and 3' GinNTalPCRfus and fused to truncated AvrXa7 variants by overlap PCR using the PCR conditions described above. Purified Gin-Avr PCR products were mixed in an equimolar ratio and digested with *SacI* and *XbaI*.

[0111] To generate designer TALEs, we used a TALEN kit (Addgene) with the following modification: pTAL1 was modified to include truncations at $\Delta 120$, $\Delta 128$, or +28. To achieve this, AvrXa7 $\Delta 120$ and AvrXa7 $\Delta 128$ fragments were PCR amplified with 5' Avr n4 or Avr n128 and 3' TalR Xba+28 and ligated into the *Bam*H1 restriction site of pTAL1 to generate pTAL $\Delta 120$ and pTAL $\Delta 128$. The plasmids pTAL $\Delta 120$ and pTAL $\Delta 128$ retained the *Esp3I* restriction sites for Golden Gate cloning. TALE arrays cloned into pTAL $\Delta 120$ and pTAL $\Delta 128$ were digested with *Bam*H1 and *XbaI* for ligation into pB-Gin-Bam.

[0112] To generate mammalian TALER expression vectors, the Gin catalytic domain was PCR amplified from pB-Gin-Avr with 5' Nhe-SD-Gin F and 3' GinGS R and ligated into the *NheI* and *Bam*HI restriction sites of pcDNA 3.1 (Invitrogen). Avr15 was digested from pTAL $\Delta 120$ or pTAL $\Delta 128$ with *Bam*H1 and *XbaI* and ligated into pcDNA-Gin-Bam to generate pcDNA-Gin-Avr expression vectors.

[0113] The pBLA substrate plasmids were constructed as previously described.

[0114] To generate pGL3 reporter plasmids, the SV40 promoter was PCR amplified from pGL3-Promoter (Promega) with the recombination site-containing primers 5' pGL3 SV40 *Bgl*III and 3' pGL3 SV40 *Hind*III and ligated into the *Bgl*III and *Hind*III restriction sites of pGL3-Promoter.

Bacterial recombination assays.

[0115] Bacterial recombination assays were performed as previously described.

Incremental truncation library.

[0116] The incremental truncation library was generated using a modified protocol previously

described. Briefly, in order to protect the Gin coding sequence from exonuclease digestion, a stuffer fragment with a *Sma*I restriction site was inserted into *Bam*H1 to generate pB-Gin-*Sma*I-Bam-Avr. This plasmid was linearized with *Nhe*I and incubated with Exonuclease III for 2.5 min at 37 °C followed by heat inactivation at 75 °C for 25 min. pB-Gin-Bam-Avr was then incubated with Klenow Fragment (3' to 5' Exo) with 200 μM dNTPs and 5 μM [α]-S-dNTPs for 30 min at 37 °C followed by heat inactivation at 80 °C for 25 min. To generate the truncation library, pB-Gin-Bam-Avr was incubated with Exonuclease III for 2.5 min at 37 °C followed by heat inactivation and subsequent blunt-ending with Mung Bean Nuclease for 1 hr at 30 °C. After digestion with *Sma*I, the blunt 3' end of the recombinase coding sequence was ligated to the blunt-ended library of TALE fragments. After transformation and purification, the plasmids were digested with *Sac*I and *Xba*I to release Gin- Δ Avr.

Mammalian reporter assays.

[0117] HEK293T cells were seeded onto 96-well plates at a density of 4×10^4 cells per well and grown in a humidified 5% CO₂ atmosphere at 37 °C. At 24 hr after seeding, cells were transfected with 150 ng pcDNA TALER expression vector, 2.5 ng pGL3 reporter plasmid, and 1 ng pRL-CMV for expression of *Renilla* luciferase using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. At 48 hr after transfection, cells were lysed with Passive Lysis Buffer (Promega) and luciferase expression was determined using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Luminescence was measured using a Veritas Microplate Luminometer (Turner Biosystems).

Results.

TALER architecture.

[0118] A quantitative system for the evaluation and directed evolution of recombinase activity has been described. In this system (Figure 1A), a GFPuv transgene flanked by recombination sites is inserted into the gene encoding TEM-1 β -lactamase. This alteration disrupts β -lactamase expression and renders *Escherichia coli* cells that harbor this plasmid (pBLA) susceptible to ampicillin. Expression of an active recombinase from the substrate-containing plasmid, however, leads to recombination between target sites and restoration of the β -lactamase reading frame. This modification establishes host-cell resistance to ampicillin and enables the isolation of active recombinase variants from the substrate-containing plasmid. By measuring the number of ampicillin-resistant transformants following plasmid purification and re-transformation, recombinase activity can be also directly assessed. Because the activity of a chimeric recombinase is dependent upon both the catalytic domain and the DBD, this split gene reassembly selection system can also be used to evaluate the effectiveness of individual DBDs. Thus, the system was adapted to determine an optimal TALER architecture.

[0119] Importantly, because the catalytic domain of the DNA invertase Gin and related serine recombinases have pre-defined catalytic specificities, TALER fusion proteins cannot be constructed using the design described for TALENs. Structural and functional studies with the $\gamma\delta$ resolvase and designed enzymes have indicated that the C-terminal E-helix mediates serine recombinase DNA recognition. In ZFRs, this helix binds DNA from the C to the N-terminus, 5' to 3'. Thus, because TALEs bind DNA in the 5' to 3' direction, it was anticipated that recombination could only occur when the TALE binding site is positioned on the opposite strand of the 20-bp core (Figure 1B).

[0120] It was chosen to generate TALERs using AvrXa7, as this TALE protein has been previously used to generate TALE nucleases and transcription factors. Conveniently, *Bam*HI restriction sites flank many TALEs, including AvrXa7 and multiple groups have used this restriction site to generate synthetic TALE fusions. Notably, this *Bam*HI fragment leaves the N-terminus of the TALE intact but removes the native effector domain from the C-terminus. This strategy was adopted and generated a Gin-AvrXa7 fusion by *Bam*H1 restriction digestion.

[0121] Gin-AvrXa7 was cloned into a pBLA selection vector containing recombination sites composed of a central 20-bp core sequence, which is recognized by the Gin catalytic domain, and two flanking 26-bp AvrXa7 binding sites. As anticipated, the Gin-AvrXa7 fusion was unable to recombine DNA when AvrXa7 binding sites were positioned adjacent to the 20-bp core (Figure 1C). However, when AvrXa7 binding sites were positioned on the opposite strand of the 20-bp core, recombination was evident (Figure 1C), indicating that recombination site orientation is a critical component for catalytic domain fusion to the TALE N-terminus. In order to further establish that N-terminal fusion is necessary for recombination, a C-terminal AvrXa7-Gin variant was constructed that contained a non-canonical fusion orientation predicted to constrain catalytic domain activity (Figure 1B and Table 5). As expected, it was determined that this C-terminal AvrXa7 fusion demonstrated negligible activity in bacterial cells (Figure 1C).

Table 5.

Variant	SEQ ID NO:	Sequence
Gin-Avr (#1) /Avr20G	195	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCATGGTTTACAG TATAAACCCCCTCCAACCAGGTGCTAA
Gin-Avr (#2)	196	TTAGCACCTGGTTGGAGGGGGTTTATA AGGTTTTGGTACCAAAATGTC TATAAACCCCCTCCAACCAGGTGCTAA
Avr-Gin (#3)	197	TATAAACCCCCTCCAACCAGGTGCTAA CTGTAAACCATGGTTTTGGA TTAGCACCTGGTTGGAGGGGGTTTATA
Avr14G	198	

Variant	SEQ ID NO:	Sequence
		TTAGCACCTGGTTGGAGGGGGTTTATA AAAACCATGGTTTA TATAAACCCCCTCCAACCAGGTGCTAA
Avr26G	199	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCCAAAACCATGGTTTACAGGGT TATAAACCCCCTCCAACCAGGTGCTAA
Avr32G	200	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCATGGTTTACAGGGTTCC
		TATAAACCCCCTCCAACCAGGTGCTAA
Avr44G	201	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCAGCTTCATCCAAAACCATGGTTTACAGGGTTCCGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr20GG	202	TTAGCACCTGGTTGGAGGGGGTTTATA TCCAAAACCGGGGTTTACAG TATAAACCCCCTCCAACCAGGTGCTAA
Avr20T	203	TTAGCACCTGGTTGGAGGGGGTTTATA CGAAATATTATAAATTATCA TATAAACCCCCTCCAACCAGGTGCTAA
Avr32GG	204	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCATCCAAAACCGGGGTTTACAGGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA
Avr32T	205	TTAGCACCTGGTTGGAGGGGGTTTATA GCTTCACGAAATATTATAAATTATCAGGTTC TATAAACCCCCTCCAACCAGGTGCTAA
Avr-G-ZF	206	GCGGGAGGCGTG TCCAAAACCATGGTTTACAGGGTTCC TATAAACCCCCTCCAACCAGGTGCTAA

Variant	SEQ ID NO:	Sequence
PthXo1-20G	207	GTGGTGTACAGTAGGGGGAGATGCA TCCAAAACCATGGTTTACAG TGCATCTCCCCCTACTGTACACCAC
PthXo1-32G	208	GTGGTGTACAGTAGGGGGAGATGCA GCTGCTTCCAAAACCATGGTTTACAGGGTGGT TGCATCTCCCCCTACTGTACACCAC

Designed truncations.

[0122] Although the Gin-AvrXa7 fusion described above catalyzed recombination, the activity of this variant was considerably lower than that of engineered ZFRs. Further, specificity analysis revealed that the Gin-AvrXa7 fusion was unable to faithfully discriminate between recognition sites containing non-cognate DBD sites and non-native 20-bp core sequences, indicating that recombination might not be Gin-mediated (Figure 1D). Recent reports have shown that TALEN activity can be enhanced when the TALE portion of the fusion protein is truncated. Thus, in order to attempt to improve TALER activity, a series of N and C-terminal AvrXa7 truncations were generated (Figure 2A).

[0123] Ten N-terminal truncations were assembled at roughly equal intervals beginning at AvrXa7 Thr 27 ($\Delta 27$) and ending at AvrXa7 Gly 268 ($\Delta 268$) (Figure 6). AvrXa7 $\Delta 150$, which has been reported as an N-terminal truncation variant for TALENs, was also generated. Two C-terminal AvrXa7 truncations were generated at positions 28 (+28) and 95 (+95). Both +28 and +95 have been reported as stable fusion points in TALENs. Each TALE truncation variant was fused to the Gin catalytic domain and this 20-member TALER library was cloned into a pBLA selection vector containing Avr-20G recognition sites. Following one round of selection in bacterial cells (Materials and Methods), individual ampicillin-resistant clones were sequenced and it was found that all selected TALERs contained either one of two N-terminal truncations: $\Delta 87$ and $\Delta 120$. Each selected clone was also +28 on the C-terminus. With the exception of a single $\Delta 120$ clone with a spontaneous 12 amino acid deletion near the fusion point ($\Delta 120^*$), the activity of these clones was quite low (Figure 2B). In this assay, Gin-based ZFRs routinely show 20-40% recombination, however, the highest activity observed amongst the selected TALER fusions was ~7% recombination (Gin-AvrXa7 $\Delta 120^*$). Because the TALE DBD is three times larger than a ZF domain (not including the required flanking peptide sequence), we reasoned that the 20-bp spacer used for these TALER constructs might not be the optimal length for recombination.

Core sequence length.

[0124] Next the effect core sequence length has on recombination was investigated by evaluating whether DNA targets containing 14 (Avr-14G), 26 (Avr-26G) and 32-bp (Avr-32G) core sites could be recombined by selected TALERs. In order to maintain the reading frame of the β -lactamase gene following recombinase-mediated reassembly, core half-sites were modified by $\pm$ 3-bps (Table 1). The 20-member TALER library described above was subjected to one round of selection against each target site variant. Although identification of TALER variants capable of recombining the shortest target was not possible, Avr-14G (data not shown), two Gin- Δ AvrXa7 variants were identified (based on the N-terminal TALE truncations Δ 87 and Δ 120 and the C-terminal truncation +28) that recombined Avr-26G and Avr-32G. In particular, clonal analysis revealed that the selected TALERs (Gin-AvrXa7 Δ 87 and Gin-AvrXa7 Δ 120) recombined DNA with longer cores (e.g., 26 and 32-bps) at least 100-fold more efficiently than shorter cores (e.g., 14 and 20-bps) (Figure 2B). Further, it was found that Gin-AvrXa7 Δ 120 recombined targets containing a cognate core sequence (Avr-26G and Avr-32G) >100-fold more efficiently than a non-cognate core (Avr-20T, Avr-20GG, Avr-32T and Avr-32GG) (Figure 2C). Interestingly, the Gin-AvrXa7 Δ 120 fusion was not as active on 44-bp cores (Avr-44G) (recombination was $\sim$ 3-fold lower than Avr-32G) (Figure 2C), indicating that core lengths between 26 and 44-bp are likely optimal for recombination by Gin-AvrXa7 Δ 120 in *E. coli*.

Incremental truncation library.

[0125] Although Gin-AvrXa7 Δ 120 showed increased recombination in comparison to Gin-AvrXa7, it was suspected that Gin-AvrXa7 Δ 120 might not be an optimal TALE fusion architecture because: (i) ZFRs containing the Gin catalytic domain recombined DNA >2-fold more efficiently than Gin-AvrXa7 Δ 120 and (ii) Gin-AvrXa7 Δ 120 was not identified from a comprehensive library of TALE truncation variants. Thus, in order to identify better fusion architectures, a screen was devised based on the generation of a library of incrementally truncated TALE DBDs.

[0126] To achieve this, a protocol was adapted as previously described to enable fusion of an unmodified N-terminal domain (Gin) to a library of truncated C-terminal fragments (AvrXa7) (Materials and Methods). N-terminal AvrXa7 truncations that spanned the region between the AvrXa7 N-terminus (Met 1) and the first AvrXa7 repeat (Leu 298) were generated by exonuclease digestion and fused to an unmodified copy of the Gin catalytic domain (theoretical number of protein variants: -300). Because previous results indicated that +28 is the optimal C-terminal truncation, we incorporated this architecture into the truncation library. TALERs were cloned into a pBLA selection vector containing Avr-32G target sites and transformed into *E. coli* ($>1 \times 10^5$ transformants). Sequence analysis confirmed an equal distribution of truncations spanning the region of interest (data not shown).

[0127] Following three rounds of selection, individual ampicillin-resistant clones were sequenced and a number of unique truncation variants were identified (Figure 3A). Consistent with the selections performed using the 20-member TALE truncation library, which suggested that the optimal N-terminal TALER fusion points were likely located in proximity to positions 87 and 120, all selected Gin-AvrXa7 variants were found to contain a truncation between positions 74 (Δ 74)

and 147 ($\Delta 147$). In particular, 26 of 73 (35.6%, $p < .001$) clones contained truncations between positions 124 ($\Delta 124$) and 129 ($\Delta 129$). From this population, truncations at position 128 ($\Delta 128$) were among the most represented.

[0128] In order to systematically determine whether selected AvrXa7 domains increased TALER activity, we evaluated the performance of isolated Gin-AvrXa7 variants against DNA substrates containing Avr-32G target sites in *E. coli*. We focused our analysis on clones containing N-terminal deletions between AvrXa7 position 92 ($\Delta 92$) and 134 ($\Delta 134$). Consistent with sequence analysis, it was found that TALERs containing N-terminal truncations between $\Delta 120$ and $\Delta 129$ recombined DNA more efficiently than variants based on comparatively longer or shorter truncations, although the $\Delta 92$ fusion was also quite active (Figure 3B). Three clones further characterized: $\Delta 74$ and $\Delta 145$ were chosen because they represented the boundaries of possible fusion points, and $\Delta 128$ was assayed because it was the most prevalent clone found in the selections. Five targets with spacer lengths from 14 to 44-bp were assayed along with three negative controls (Avr32T, Avr32GG, and PthXo1-32G). It was determined that Gin-Avr32G $\Delta 74$ and Gin-Avr32G $\Delta 145$ had modest activity on spacers longer than 20-bp, whereas Gin-Avr32G $\Delta 128$ recombined DNA with efficiencies comparable to the ZFR GinC4 (Figure 3C). Furthermore, specificity analysis revealed that Gin-Avr32G $\Delta 74$, Gin-Avr32G $\Delta 128$, and Gin-Avr32G $\Delta 145$ could recombine substrates harboring cognate cores >100-fold more efficiently than non-cognate cores (Avr-32T, Avr-32GG and PthXo1-32G) (Figure 3C). Together, these results suggest that TALE proteins containing N-terminal deletions between $\Delta 120$ and $\Delta 129$ represent an optimal truncation for fusion to a recombinase.

Incorporation of synthetic TALE repeat arrays.

[0129] The studies described above used the native DBDs of the naturally occurring AvrXa7 TALE protein. In order to determine whether designed TALE repeat arrays can be incorporated into the selected Gin- Δ AvrXa7 frameworks, a series of synthetic TALE proteins (15 to 20 repeats in length) were generated designed to target the AvrXa7 binding site (Figure 7). TALE proteins were constructed using a publicly available TALEN plasmid set (Addgene). The cloning plasmid was modified to include the +28 C-terminal truncation and either the $\Delta 120$ or $\Delta 128$ N-terminal truncation. Designed TALEs were fused to the Gin catalytic domain (denoted as Gin-Avr15 $\Delta 120$ and Gin-Avr15 $\Delta 128$) and cloned into a pBLA selection vector containing Avr-32G or Avr-32T target sites.

[0130] Activity analysis in *E. coli* revealed that both Gin-Avr15 $\Delta 120$ and Gin-Avr15 $\Delta 128$ could be used to recombine DNA when fused to an active catalytic domain and that incorporation of synthetic repeats provided an increase in activity (Figure 4A). Importantly, each TALER displayed stringent selectivity, recombining target sites that contained cognate cores >1,000-fold more efficiently than non-cognate cores (Figure 4B). Surprisingly, TALERs based on the $\Delta 120$ truncation were also found to recombine DNA as effectively as TALEs based on the $\Delta 128$ architecture (Figure 4A), indicating that designed TALEs may be less sensitive to N-terminal truncation than those containing the native AvrXa7 DBD.

[0131] To further demonstrate that the TALER architecture described herein can be reprogrammed to target any DNA sequence, a synthetic enzyme was created designed to target the sequence recognized by the naturally occurring TALE protein PthXo1 (Gin-Pth15 Δ 120). It was found that Gin-Pth15 Δ 120 was highly active on its cognate substrate and that both Gin-Pth15 Δ 120 and Gin-Avr15 Δ 120 showed a >600 fold increase in recombination for targets with their cognate binding sites (Figure 4A). The activity of a series of designed TALERS containing DBDs between 15 and 20 repeats in length was also assessed and found that each fusion catalyzed recombination with similarly high efficiency and specificity (Figure 4B), demonstrating that chimeric recombinases that incorporate synthetic TALE repeat arrays can be used for site-specific recombination.

TALER activity in mammalian cells.

[0132] It was also determined whether TALERS could modify DNA in mammalian cells. To achieve this, we used an episomal reporter assay that enables rapid assessment of recombinase activity in cell culture. In this assay, human embryonic kidney (HEK) 293T cells are co-transfected with a recombinase expression vector and a reporter plasmid (pGL3) that contains a luciferase gene under the control of a SV40 promoter flanked by recombination sites. Transient expression of the appropriate recombinase leads to excision of the SV40 promoter and reduced luciferase expression in cells. Recombinase activity is thus directly proportional to the fold-reduction in luciferase expression.

[0133] Co-transfection of Gin-Avr15 Δ 120 with a reporter plasmid harboring Avr-44G recognition sites (pGL3-Avr-44G) led to a ~20-fold reduction in luciferase expression as compared to transfection of pGL3-Avr-44G alone (Figure 5A). Despite the fact that Gin-Avr15 Δ 120 showed similar activity to the ZFR GinC4 in *E. coli*, we found that GinC4 reduced luciferase expression by >80-fold after co-transfection with its cognate target plasmid, pGL3-C4-20G (Figure 5A). This discrepancy may be due to the comparatively shorter intervening DNA sequence between recombinase target sites in pGL3 than pBLA or differential expression between TALERS and ZFRs in mammalian cells. The underlying cause for this disparity, however, remains unclear. Finally, although 32-bp was determined to be the optimal core sequence length for TALERS in *E. coli*, it was determined that co-transfection of Gin-Avr15 Δ 120 with pGL3-Avr-32G led to only a 6-fold reduction in luciferase expression (Figure 5A). The underlying cause behind this disparity also remains unclear.

[0134] Next whether a ZFR (GinC4) and a TALER (Gin-Avr15 Δ 120) could form a compatible heterodimer in mammalian cells was investigated. To evaluate this possibility, a hybrid recombination site was generated in which the AvrXa7 binding site and the C4 zinc-finger binding site (GCG GGA GGC GTG; SEQ ID NO: 279) flank the core sequence recognized by the Gin catalytic domain (pGL3-Avr-G-ZF) (see Table 2). Surprisingly, co-transfection of pGL3-Avr-G-ZF with GinC4 and Gin-Avr15 Δ 120 led to a >140-fold reduction in luciferase expression as compared to pGL3-Avr-G-ZF (Figure 5B), whereas transfection with either GinC4 or Gin-Avr15 Δ 120 with pGL3-Avr-G-ZF led to a negligible decrease in reporter gene expression. These results demonstrate that generating ZF-TALE heterodimers represents a potentially effective approach

for improving the targeting capacity of chimeric recombinases.

Discussion.

[0135] Unlike ZFPs, which contain a very minimal fusion architecture, TALE DBDs require native protein framework on either side of the DBD array to function. The so-called 0th and 1st repeats, which mediate binding of the thymidine residue at position 0 and are found in almost all known TALE recognition sites, represent such an N-terminal framework. A recent crystal structure provided a description of the binding of the position 0 thymine, yet there remains insufficient data to determine a minimal TALE architecture. Indeed, all studies to date have used an N-terminal truncation containing considerably more residues than those required to mediate binding at position 0. It remains uncertain what role this part of the protein has in enabling the proper DNA binding conformation or what might constitute a minimal TALE domain. Although initial attempts to generate functional TALE chimeras were based on fusion to full-length TALE proteins, more recent studies have focused on the identification of unique C-terminal truncations that improve effector domain function in the context of the $\Delta 150$ N-terminal architecture. A previous report indicated that deletion of N-terminal residues 2-153 ($\Delta 150$) of the AvrBs3 TALE removes the domain required for translocation of the TALE from its native bacteria to the target plant cell but does not compromise transcription factor activity.

[0136] Developing an active TALER, however, necessitated that unique N-terminal TALE variants be identified. A broad, systematic survey was initially conducted of N-terminal TALEs with the C-terminal truncations +28 and +95 and found that only two domains ($\Delta 87$ with +28 and $\Delta 120$ with +28) demonstrated sufficiently high activity for further analysis. A secondary analysis based on incremental truncation of the AvrXa7 N-terminus led to the identification of a broad cluster of truncation variants centered between AvrXa7 position 74 ($\Delta 74$) and position 145 ($\Delta 145$). Of the clones recovered in this experiment, 38% contained truncations between positions $\Delta 119$ and $\Delta 128$, and a survey of data obtained on TALERs with fusions in this region showed high activity. In particular, it was determined that TALERs based on N-terminal truncations from this region ($\Delta 128$ and $\Delta 120$) could be used to recombine DNA in bacteria and mammalian cells. The clustering of truncation variants between $\Delta 119$ and $\Delta 128$ may also be indicative of the intrinsic stability of this region.

[0137] ZFRs typically catalyze recombination between target sites 44 to 50-bp in length. Each target site contains a central 20-bp core sequence, which is recognized by the recombinase catalytic domain, and two adjacent ZFP binding sites. The fusion orientation of TALERs, however, necessitates that TALE binding sites are on the opposite strand relative to the central core sequence. This unique geometry led us to investigate the minimum core sequence requirements for recombination. Because of the length of TALE DBDs (TALE repeats are 3 to 4 times longer than ZFPs) and the extended N-terminal linker between the catalytic domain and the TALE domain, we reasoned that longer core sequences (32 or 44-bp) would be necessary for recombination. Indeed, with the exception of a TALE variant harboring a spontaneous deletion ($\Delta 120^*$), most N-terminal truncation variants identified in this study demonstrated optimal performance against 32-bp cores. These results are consistent with those reported with TALENs,

which unlike ZFNs require significantly longer spacer sequences (e.g. TALENs: 17 to 20-bp, ZFNs: 5 to 6-bp) to efficiently cleave DNA. In support of these observations, it was found that selection for unique N-terminal truncation variants against a short core sequence (14-bp) did not yield any clones.

[0138] Gin-AvrXa7 Δ 128 was identified as an optimal TALE fusion, but subsequent studies using synthetic TALE proteins generated using a publicly available TALE assembly kit indicated that Δ 128 and Δ 120-based TALERs showed similar activity in *E. coli*. These designed TALEs were based on a chimeric protein derived from the closely related and naturally occurring Tal1c and PthXo1 TALE proteins. Although TALEs share high homology, they are not identical. While polymorphisms in RVD repeats outside of residues 12 and 13 have been shown to have no effect on TALE fusion activities, to our knowledge no systematic evaluation of differences in TALE framework outside the DBDs has been reported. As demonstrated by the analysis of the incremental truncation library, minor amino acid alterations can significantly influence the activity of a particular fusion. Thus, some of the discrepancy in activity we observed between Gin-AvrXa7 Δ 120 and the synthetic Gin-Avr15 Δ 120 may be attributable to the sequence variations between AvrXa7 framework and the TALE framework architecture used previously.

[0139] The four RVDs (NI: A, HD: C, NG: T, and NN: G) favored for construction of synthetic TALEs are the most prevalent in nature; however, it remains to be determined whether these repeats represent the most specific RVD modules. For the 26-repeat AvrXa7 TALE, a synthetic version targeting the same sequence would have 16 changes in RVD composition (Figure 7). It was hypothesized that because they are more commonly found in nature, the four RVDs selected for synthetic use might have a higher affinity for their cognate bases than other RVDs. If this were the case, it would be reasonable to assume that a TALE created with the synthetic RVD repeats could have higher DNA-binding affinity than a TALE using the native domains. Although the issue of RVD affinity was not directly addressed, it was determined that TALERs containing synthetic repeat arrays were more active than constructs, which contained the native AvrXa7 DBD. TALERs with synthetic DBDs showed approximately two-fold higher activities than constructs containing the native repeats, despite containing significantly fewer DBDs. Additionally, the gain in activity observed with the synthetic arrays was not correlated with any increase in off-target recombination.

[0140] Several studies have shown that TALEs can tolerate some mismatches in their target sequence. These findings are unsurprising, as RVDs that are positively associated with particular bases have been shown to tolerate non-cognate bases in nature. The cooperative specificity afforded by TALERs could be used to circumvent potential limitations, however. Because the catalytic domain contributes specificity to recombination, it is envisioned that designer TALERs capable of selectively modifying highly homologous genomic sequences could be generated as well. Indeed, it has been recently demonstrated that recombinase catalytic specificity can be effectively reprogrammed to target unnatural core sites.

EXAMPLE 2

SELECTION OF NOVEL 0TH RESIDUE SPECIFICITY

[0141] A new class of Tal-based DNA binding proteins was engineered. TAL (transcription activator-like) effectors constitute a novel class of DNA-binding proteins with predictable specificity. Tal effectors are employed by Gram-negative plant-pathogenic bacteria of the genus *Xanthomonas* which translocate a cocktail of different effector proteins via a type III secretion system (T3SS) into plant cells where they serve as virulence determinants. DNA-binding specificity of TALs is determined by a central domain of tandem repeats. Each repeat confers recognition of one base pair (bp) in the DNA. Rearrangement of repeat modules allows design of proteins with desired DNA-binding specificities with certain important limitations. For example, the most constraining feature of targeting a DNA sequence with a Tal domain is the requirement that the Tal DNA site start with the base T and sometimes C. Targeting a binding site starting with a G or A base has not been possible at the -1 position. Tal-recombinase activity selections were used to select for Tal DNA binding domains that lack this restriction by targeting mutations to the -1 and 0th RVD regions. The practical consequences of this discovery are vast since now every DNA sequence can be targeted with new Tal domains facilitating new unrestricted approaches to TAL transcription factors to turn transcription on/up or off/down, to target TAL nucleases to knock out gene function or to direct homologous recombination or to target our own TAL recombinases or other TAL enzymes.

[0142] For G specificity at the (-1) position, the amino acids QWSG (SEQ ID NO: 209) were first randomized using an NNK codon strategy within the (-1) domain of the GinAvr15Δ128-synthetic protein. Following 3 rounds of tal recombinase activity selection of the resulting library, novel tal binding domains with the selected sequences RSWG (SEQ ID NO: 210) and SRSG (SEQ ID NO: 211) in the targeted region were selected. These were then shown to bind G at the 0th position of the target sequence over the parental T recognized by the starting clone. The selection was repeated randomizing the KQW region shown below in red that overlaps with the QWSG (SEQ ID NO: 212) selected initially. Now clones with selected SSR, SRA, SRC, and KRC sequences were selected. All selected Tal binding domains were assayed in binding studies to defined oligos bearing the G substitution and shown to now preferentially bind the sequence G-ATAAACCCCCTCCAA (SEQ ID NO: 213). Note that the Tal recombinase activity selection was performed using this same sequence. The starting Tal binding protein the GinAvr15Δ128 binds T-ATAAACCCCCTCCAA (SEQ ID NO: 214). Subsequence testing of Tal nucleases bearing the selected mutations verify the G specify of these sequences allowing for this novel class of Tals to be developed for the first time. Selected sequences are portable to Tals derived from other species.

Table 6.

Selections	
SEQ ID NO: 215	ATHEDIVGVGKQWSGARALEALLTDAGELRGPPPLQ (-1 domain)
SEQ ID NO: 216	ATHEDIVGVGK QWSG ARALEALLTDAGELRGPPPLQ (randomized AA in bold)
SEQ ID NO: 217	KQWSG-starting clone sequence

Selections	
SEQ ID NO: 218	<i>KRSNG-selected to bind G</i>
SEQ ID NO: 219	<i>KSRSNG-selected to bind G</i>
SEQ ID NO: 220	ATHEDIVGVG KQWS GARALEALLTDAGELRGPPLQ
SEQ ID NO: 221	KQWS G-WT
	<i>SSR-selected to bind G</i>
	<i>SRA-selected to bind G</i>
	<i>SRC-selected to bind G</i>
	<i>KRC-selected to bind G</i>

[0143] Selections were also performed using this same library to target A. In this study, sequences PRG, PTR, and PKD were selected. All selected Tal binding domains were assayed in binding studies to defined oligos bearing the A substitution and shown to now preferentially bind the sequence A-ATAAACCCCCTCCAA (SEQ ID NO: 222). Note that the Tal recombinase activity selection was performed using this same sequence. The starting Tal binding protein the GinAvr15Δ128 binds T-ATAAACCCCCTCCAA (SEQ ID NO: 223). Subsequence testing of Tal nucleases bearing the selected mutations verify the A specificity of these sequences allowing for this novel class of Tals to be developed for the first time. Subsequent refinements in binding activities can be achieved by random mutagenesis of the N-terminal domain or target mutagenesis of the KRGG (SEQ ID NO: 224) sequence within the Oth domain and reselection in the recombinase system.

EXAMPLE 3

SELECTIONS

[0144] For context dependent RVD selections and selections of RVDs with new specificities, libraries were created that randomize the HD sequence emboldened below. LTPDQVVAIASH**HDGG**KQALETVQRLLPVLCQDHG (prototype RVD sequence; SEQ ID NO: 225)

[0145] Typically the library allows all amino acids at these two positions, though libraries limited to N, D, H, K, and Q amino acids are often successful substitutes for the H residue. Alternatively larger libraries that randomized the **SHDG** (SEQ ID NO: 226) and **ASHDG** (SEQ ID NO: 227) regions allow for the selection of unique RVD specificities with context dependent characteristics.

[0146] Tal recombinase activity selections then rapidly allow for the selection of new specificities within the targeted RVD domain. The resulting RVDs can be highly modular or context dependent in their sequence recognition and can be then used to create Tal nucleases and transcription

factors.

[0147] Utility of this technology includes unrestricted approaches to TAL transcription factors to turn transcription on/up or off/down, to target TAL nucleases to knock out gene function or to direct homologous recombination or to target our own TAL recombinases or other TAL enzymes for use as tools and therapeutics.

[0148] Advantages and the practical consequences of this discovery are vast since now every DNA sequence can be targeted with our new Tal domains and their specificities can be readily optimized.

EXAMPLE 4

DIRECTED EVOLUTION OF TALE N-TERMINAL DOMAIN TO ACCOMMODATE 5' BASES OTHER THAN THYMINE

[0149] Transcription activator-like effector (TALE) proteins can be designed to bind virtually any DNA sequence of interest. The DNA binding sites for natural TALE transcription factors (TALE-TFs) that target plant avirulence genes have a 5' thymidine. Synthetic TALE-TFs also have this requirement. Recent structural data indicate that there is an interaction between the N-terminal domain (NTD) and a 5' T of the target sequence. A survey of the recent TALE nuclease (TALEN) literature yielded conflicting data regarding the importance of the first base of the target sequence, the N₀ residue. Additionally, there have been no studies regarding the impact of the N₀ base on the activities of TALE recombinases (TALE-Rs). Here, the impact of the N₀ base is quantified in the binding regions of TALE-Rs, TALE-TFs, TALE DNA-binding domains expressed as fusions with maltose binding protein (MBP-TALEs) and TALENs. Each of these TALE platforms have distinct N-and C-terminal architectures, but all demonstrated highest activity when the N₀ residue was a thymidine. To simplify the rules for constructing effective TALEs in these platforms, and allow precision genome engineering applications at any arbitrary DNA sequence, we devised a structure-guided activity selection using our recently developed TALE-R system. Novel NTD sequences were identified that provided highly active and selective TALE-R activity on TALE binding sites with 5' G, and additional domain sequences were selected that permitted general targeting of any 5' N₀ residue. These domains were imported into TALE-TF, MBP-TALE and TALEN architectures and consistently exhibited greater activity than did the wild-type NTD on target sequences with non-T 5' residues. The novel NTDs are compatible with the golden gate TALEN assembly protocol and now make possible the efficient construction of TALE transcription factors, recombinases, nucleases and DNA-binding proteins that recognize any DNA sequence allowing for precise and unconstrained positioning of TALE-based proteins on DNA without regard to the 5' T rule that limits most natural TALE proteins.

[0150] The following Material and Methods were utilized in this Example.

Oligonucleotides.

[0151] Primers and other oligonucleotides (Table 4 below) were ordered from Integrated DNA Technologies (San Diego, CA).

Table 7. Primers.

Primer	SEQ ID NO:	Sequence
KXXG Lib Rev	228	TCTCAACTCCCCCGCCTCCGTGAGCAAGGCCTCCAGAGCGCGTGCC CCMNNMNNTTTGCCGACGCCAACGATGTCTTCGTG
KXXXX Lib Rev	229	TCT CAA CTC CCC CGC CTC CGT GAG CAA GGC CTC CAG AGC GCG TGC MNN MNN MNN MNN TTT GCC GAC GCC AAC GAT GTC TTC GTG
XXXSG Lib Rev	230	CCCGCCTCCGTGAGCAAGGCCTCCAGGGCGCGTGCGCCGGAMNNM NNMNNGCCGACGCCAACGATGTCTTCGTGTGTGCGC
KRGG Lib Rev	231	GGC ACC CGT CAG TGC ATT GCG CCA TGC ATG CAC TGC CTC CAC TGC GGT CAC MNN MNN MNN MNN TGC AAT CTT GAG AAG TTG GCC TGT GTC
Goldy TALEN fwd	232	AGAGAGAGAAGAAAATGAGATCTCCTAAGAAAAAGAGGAAGGTGC AGGTGGATCTACGCACGCTCGGCTAC
NTD-dHax3 Fwd	233	AGGAAGAAGAGAAGCATGAGATCTCCTAAGAAAAAGAGGAAGGTG ATGGTGGACTTGAGGACACTCGGTTA
NTD-dHax3 Rev	234	AAGAGAAGAAGAAGAAGCATTGCGCCATGCATGCACTGCCTCTA
pTa1127 Not1 fwd	235	CCC GCC ACC CAC CGT GC
N-Term Sph1	236	TGC TCT ATG CAT GCA CTG CCT CC
pTAL127-SFI Fwd	237	AGA GAA GAG AAG AGA AGG CGC CCG CGG CCC AGG CGG CCT CGG GAT CCC CTC GGC CTC CGC GCG CCA AG

Primer	SEQ ID NO:	Sequence
pTAL127-SFI +95 Rev	238	AGA GAG AGA GAG AGA GTC TAG AGG CCG GCC TGG CCG CTC ATC CCG AAC TGC GTC ATG GCC TCA TC
pTAL127 Xba +28 Rev	239	GCC CCA GAT CCT GGT ACG CTC TAG AGG
Avr 5'A biotin hairpin	240	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATTGG GTT TTC CCAAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'T biotin hairpin	241	5'Biosg/ATC TTA GCA CCT GGT TGG AGG GGG TTT ATAGG GTT TTC CCTAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'G biotin hairpin	242	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATCGG GTT TTC CCGAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
Avr 5'C biotin hairpin	243	5'BiosgATC TTA GCA CCT GGT TGG AGG GGG TTT ATGGG GTT TTC CCCAT AAA CCC CCT CCA ACC AGG TGC TAA GAT
CCR5-inner fwd	244	TTAAAAGCCAGGACGGTCAC
CCR5-inner rev	245	TGTAGGGAGCCCAGAAGAGA
CCR5-outer fwd	246	ACAGTTTGCATTCATGGAGGGC
CCR5-outer rev	247	CCGAGCGAGCAAGCTCAGTT
CCR5-indel fwd	248	CGCGGATCCCCGCCCAAGTGGGACTTTG
CCR5-indel rev2	249	CCGGAA TTCACCTGTT AGAGCT ACTGC
pGL3 NTD stuffer fwd	250	AGA GAG AGA GAG AGG CGG CCG CCC TAC CAG GGA TTT CAG TCG ATG TAC ACG TTC
pGL3 NTD stuffer rev	251	AAG AAG AAG AAG GAA GAG AAG TAG GCC TGT CAT CGT CGG GAA GAC CTG CGA CAC CTG C
	252	

Primer	SEQ ID NO:	Sequence
pg13 5X Avr XhoI		ACTGCTATCCGAGTATAAAACCCCTCCAACCAGGTATAAAACCCCT CCAACCAGGTATAAAACCCCTCCAACCAGGTATAAAACCCCTCCAA CCAGGTATAAAACCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
AvrXa7 32G A F	253	TTAATTAAGAGTCTAGAttageacctgggtggaggggttatTgetteaTCCAAAACC ATGGTTTACAGggttccAATAAACCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G A R	254	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATTgc aaccCTGTAAACCATGGTTTTGGAtgaagcAATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G G F	255	TTAATTAAGAGTCTAGAttageacctgggtggaggggttatCgetteaTCCAAAACC ATGGTTTACAGggttccGATAAACCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G G R	256	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATCgc aaccCTGTAAACCATGGTTTTGGAtgaagcGATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC
AvrXa7 32G C F	257	TTAATTAAGAGTCTAGAttageacctgggtggaggggttatGgetteaTCCAAAACC ATGGTTTACAGggttccCATAAACCCCTCCAACCAGGTGCTAAAGAT CTAGGAGGAATTTAAAATGAG
AvrXa7 32G C R	258	ACTGACCTAGAGAAGCTTTTAGCACCTGGTTGGAGGGGGTTTATGgc aaccCTGTAAACCATGGTTTTGGAtgaagcCATAAACCCCTCCAACCAG GTGCTAACTGCAGTTATTTGTACAGTTCATC

Luciferase. Vector = pgl3 basic. XhoI/SphI		
Forward target containing:	SEQ ID NO:	
5x Avr15 n-1c xhoF:	259	actgctatctcgagcTATAAACCCCCCTCCAACCAGGcTATAAACCCCCCTCCAACCAGGcTATAAACCCCCCTCCAACCAGGcTATAAACCCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=A n-lc	260	actgctatctcgagcAATAAACCCCCCTCCAACCAGGcAATAAACCCCCCTCCAACCAGGcAATAAACCCCCCTCCAACCAGGcAATAAACCCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=C n-lc	261	actgctatctcgagcCATAAACCCCCCTCCAACCAGGcCATAAACCCCCCTCCAACCAGGcCATAAACCCCCCTCCAACCAGGcCATAAACCCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
5x Avr15 0=G n-lc	262	actgctatctcgagcGATAAACCCCCCTCCAACCAGGcGATAAACCCCCCTCCAACCAGGcGATAAACCCCCCTCCAACCAGGcGATAAACCCCCCTCCAACCAGGATCTGCGATCTAAGTAAGCT
Luciferase Reverse Primer:	263	TCAGAAACAGCTCTTCTTCAAATCT

Generation of TALE-R NTD evolution plasmids.

[0152] The TALE-R system previously reported was adapted for this study. Briefly, pBCS (containing chloramphenicol and carbenicillin resistance genes) was digested with HindIII/Spe1. The stuffer (Avr X, where X is the N₀ base), containing twin recombinase sites, was digested with HindIII/Xba1 and ligated into the vector to create a split *beta-lactamase* gene. pBCS AvrX was then digested with BamH1/Sac1, and Gin127-N-stuffer-Avr15 was digested with BamH1/Sac1 and ligated into the vector to create Gin127-N-stuffer-Avr15-X. The stuffer was digested with Not1/Stu1 for evolutions at the N₋₁ TALE hairpin and Not1/Sph1 for evolutions at the N₀ TALE hairpin.

Generation of TALE NTD evolution libraries.

[0153] Primer ptal127 Not1 fwd and reverse primers KXXG lib rev or KXXXX lib rev were used to

generate N-terminal variants at the N₁ TALE hairpin and were subsequently digested with Not1/Stu1 then ligated into digested Gin127-AvrX. Forward primer ptal127 Not1 fwd and reverse primer KRGG Lib Rev were used to PCR amplify a library with mutations in the N₀ TALE hairpin. This was subsequently digested with Not1/Sph1 and ligated into Not1/Sph1-digested Gin127-AvrX.

TALE-R NTD evolution assay.

[0154] Round 1 ligations were ethanol precipitated and transformed into electrocompetent Top10 F' cells then recovered in SOC for 1 h. The cells were grown overnight in 100 ml Super Broth (SB) media containing 100 mg/ml chloramphenicol. DNA was isolated via standard procedures. The resulting plasmid DNA (Rd 1 input) was transformed into electrocompetent Top10F' cells; cells were grown overnight in 100 ml of SB containing 100 mg/ml carbenicillin and 100 mg/ml chloramphenicol. Plasmid DNA was isolated via standard procedures. Round 1 output was digested with Not1/Xba1 and ligated into the Gin127-AvrX vector with complementary sticky ends. This protocol was repeated three to four times when a consensus sequence was observed and clones were characterized.

Measurement of N-terminal TALEN activity.

[0155] Four TALEN pairs containing each possible base were generated using the golden gate protocol. Fusion A and B plasmids were directly ligated via second golden gate reaction into the Goldy TALEN (N Δ 152/C +63) framework. The NTD was modified by digesting the pCAG vector with BglII/Nsi1 and ligating with PCR amplified NTD digested with BglII/Nsi1. TALEN pairs (50-75 ng each TALEN/well) were transfected into HeLa cells in wells of 96-well plates at a density of 1.5×10^4 cells/well. After transfection, cells were placed in a 37° C incubator for 24 h, then were moved to 30° C for 2 days and then moved to 37° C for 24 h. Genomic DNA was isolated according to a published protocol, and DNA mutation rates were quantified with the Cell Surveyor assay and by sequencing. For Cell assays, genomic DNA was amplified by nested PCR, first with primers CCR5 outer fwd/CCR5 outer rev and then with CCR5 inner fwd/CCR5 inner rev. For sequencing of indels, the second PCR was performed with CCR5 indel fwd/CCR5 indel rev. Fragments were then digested with BamH1/EcoR1 and ligated into pUC19 with complementary digestion.

TALE-TFs and luciferase assay.

[0156] Variant NTDs from the recombinase selection were PCR amplified with primers ptal127 SFI fwd and N-Term Sph1. The PCR product was amplified and digested with Not1/Stu1 and ligated into pTAL127-SFI Avr15, which contains twin SFI-1 digestion sites facilitating transfer of the N-terminal-modified TALE from pTAL127-SFI Avr15 into pcDNA 3.0 VP64. Corresponding TALE binding sites were cloned into the pGL3 Basic vector (Promega) upstream of the luciferase

gene. For each assay, 100 ng of pcDNA was co-transfected with 5 ng of pGL3 vector and 1 ng of pRL *Renilla* luciferase control vector into HEK293t cells in a well of a 96-well plate using Lipofectimine 2000 (Life Technology) according to manufacturer's specifications. After 48 h, cells were washed, lysed and luciferase activity assessed with the Dual-Luciferase reporter system (Promega) on a Veritas Microplate luminometer (Turner Biosystems). Transfections were done in triplicate and results averaged.

MBP-TALE assay.

[0157] Affinity assays of MBP-TALE binding to biotinylated oligonucleotides were performed using a protocol previously described. Briefly, AvrXa7 TALE domains were expressed from pMAL MBP-AvrXa7 plasmid in XL1-Blue cells and purified on amylose resin. Biotinylated oligonucleotides containing the target AvrXa7 target site with modified residues were used to determine TALE-binding activity in sandwich enzyme-linked immunosorbent assay format. Antibodies targeting the MBP substituent were used for assay development.

Results.

Preliminary analysis of the 5' T rule.

[0158] A recent crystal structure of a TALE protein bound to PthXo7 DNA sequence revealed a unique interaction between W232 in the N-1 hairpin with a thymidine at the 5' end of the contacted region of the DNA substrate (the N₀ base). This study provided a structural basis for the previously established 5' T rule reported when the TALE code was first deciphered (Figure 18A and B). There are conflicting data regarding the importance of the first base of the target sequence of TALENs. The requirement for a 5' T in the target DNA was initially assessed in the context of TALE-Rs using four split *beta lactamase* TALE recombinase selection vectors containing four AvrXa7 binding sites with all possible 5' residues flanking a Gin32G core (Figure 18C). Recognition of the N₀ residue by TALE-TFs was then evaluated using four luciferase reporter vectors containing a pentamer AvrXa7 promoter region with recognition sites containing each possible 5' residue (Figure 18D). With bases other than a 5' T, we observed decreases in activity up to >100-fold in TALE-Rs and 1000-fold in TALE-TFs relative to the sequence with a 5' T (Figure 18C and D). These reductions were observed despite variations in the C-terminal architectures of these chimeras that reportedly remove the 5' T bias, especially in the presence of a greatly shortened C-terminal domain (CTD). Enzyme-linked immunosorbent assay also indicated decreased affinity of MBP-TALE DNA-binding proteins toward target oligonucleotides with non-T 5' residues (Figure 18E). Finally, examination of the activity of designed TALENs with wild-type NTDs on targets with non-T 5' nucleotides showed up to 10-fold decrease in activity versus those with a 5' T (Figure 18F). The results indicate that a 5' T is an important design parameter for maximally effective TALE domains in the context of recombinases, transcription factors, nucleases and simple DNA-binding proteins.

Evolution of the TALE NTD to accommodate non-T 5' residues.

[0159] To create a more flexible system for DNA recognition, it was hypothesized that the recently developed TALE-R selection system could be utilized to evolve the NTD of the TALE to remove the 5' T constraint (Figure 23). Libraries were generated with residues K230 through G234 randomized, and TALE-Rs with activity against each possible 5' base were isolated after several rounds of selection (Figure 19A-C). The most active selected clones exhibited strong conservation of K230 and G234; the former may contact the DNA phosphate backbone, and the latter may influence hairpin loop formation (Figure 24). In the case of library K230-W232, K230S was frequently observed but had much lower activity than K230R or K230 variants in nearly all variants assayed individually. One clone (NT-G) of several observed with a W232 to R232 mutation demonstrated a significant shift of selectivity from 5' T to 5' G; the sequence resembles that of the NTD of a recently described *Ralstonia* TALE protein in this region. The *Ralstonia* NTD, in the context of plant transcription factor reporter gene regulation, has been reported to prefer a 5' G in its substrate (see Figure 25 for a protein alignment). Residue R232 may contact the G base specifically, as indicated by the stringency of NT-G for 5' G. The preference of NT-G for a 5' G was comparable with the specificity of the wild-type domain for 5' T. NTD variants specific for 5' A or 5' C were not able to be derived, but a permissive NTD, NT- α N, was obtained that resembles the K265-G268 N₀ hairpin that accepts substrates with any 5' residue and maintains high activity. It was hypothesized that this variant makes enhanced non-specific contacts with the DNA phosphate backbone compared with the wild-type NTD, enhancing the overall binding of the TALE-DNA complex without contacting a specific 5' residue. It was hypothesized that a shortened hairpin structure would allow selection of variants with specificity for 5' A or 5' C residues. A library with randomization at Q231-W232 and with residue 233 deleted was designed to shorten the putative DNA-binding loop. Recombinase selection revealed a highly conserved Q231Y mutation that had high activity in a number of clones (Figure 19D). In particular, NT- β N demonstrated improved activity on substrates with 5' A, C and G but diminished activity on 5' T substrates compared with TALEs with the wild-type NTD (Figure 19E).

Applications of evolved TALE NTDs.

[0160] To assess the portability of the evolved NTDs in designer TALE fusion protein applications, optimized NTDs were incorporated into TALE-TFs, MBP-TALEs and TALENs. TALE-TFs with NT-G, NT- α N and NT- β N domains demonstrated 400-1500-fold increases in transcriptional activation of a luciferase target gene bearing operator sites without a 5' T residue when compared with the TALE-TF with the NT-T domain. The NT-G-based TF retained the 5' G selectivity as observed in the TALE-R selection system. The activities of NT- α N- and NT- β N-based TFs against all 5' nucleotides tracked the relative activity observed in the recombinase format (Figure 20). MBP-TALEs also exhibited greater relative binding affinity for target oligonucleotides with sites that did not have a 5' T than did the wild-type MBP-TALE (Figure 26), providing further evidence that the selected domains enhanced recognition of or tolerance for non-thymine 5' bases.

[0161] Four of the optimized NTDs were then imported into the Goldy TALEN framework. For these experiments, four substrates were constructed within the context of the $\Delta 32$ locus of the CCR5 gene (Figure 21A). Each substrate contained a different 5' residue. Experiments included TALENs with wild-type (NT-T) and dHax3 NTDs (dHax3 is commonly used NTD variant isolated from *Xanthomonas campestris*) with specificity for 5' T, to benchmark gene editing activity. The substrate TALEN pairs were designed to retain as much RVD homology (50-90%) as possible to determine the activity enhancing contributions of the variant NTDs (Figure 21A).

[0162] Activities of the TALENs were analyzed both by sequencing and by using the Cell assay. The selected domains exhibited increases in gene editing activity between 2- and 9-fold for the non-T 5' residues when compared with activities of the TALEN containing the wild-type domain (Figure 21 and Figure 27). Activity was highest on TALEN pair T1/T2 with wild-type or dHax3 NTD. The TALEN pair substrate G1/G2 was processed most effectively by TALENs with NT- α N, NT- β N and NT-G, with 2.0-3.5-fold enhancement versus NT-T. NT- α N had activity 9- and 2-fold higher than the wild-type NT-T on TALEN pairs A1/A2 and C1/C2, respectively. Although the impact of a mismatch at the 5' residue is more modest in TALENs than in TALE-TF and TALE-R frameworks, the optimized NTDs greatly improved TALEN activity when used in gene editing experiments.

Discussion.

[0163] Most, but not all, previous studies have suggested that a thymidine is required as the 5'-most residue in design of optimal TALE DNA-binding domains. The analyses described here indicate that a thymidine is optimal, and in some cases critical, for building functional TALE fusion proteins. This requirement therefore imposes limitations on the sequences that can be effectively targeted with TALE transcription factor, nuclease and recombinase chimeras. Although this requirement theoretically imposes minor limitations on the use of TALENs for inducing gene knockout, given their broad spacer region tolerance, NTD's that can accommodate any 5' residue would further simplify the rules for effective TALE construction and greatly enhance applications requiring precise TALE placement for genome engineering and interrogation (e.g. precise cleavage of DNA at a defined base pair using TALENs, seamless gene insertion and exchange via TALE-Recombinases, displacement of natural DNA-binding proteins from specific endogenous DNA sequences to interrogate their functional role, the development of orthogonal transcription factors for pathway engineering, the synergistic activation of natural and synthetic genes wherein transcription factor placement is key and many other applications). Other uses in DNA-based nanotechnology include decorating DNA nanostructures/origami with specific DNA-binding proteins. Here, targeting to specific sites is constrained based on DNA folding/structure and thus being able to bind any site is critical. Elaboration of these structures and devices with DNA-binding proteins could be a fascinating approach to expanding function. Indeed, it is not difficult to imagine many applications for DNA binding proteins and their fusions when all targeting constraints are removed. Encouraged by these potential applications, we aimed to develop NTDs that enable targeting of sites initiated at any base.

[0164] The recently developed TALE-R system was used to evolve the NTD of the TALE to

remove the 5'-T constraint. In three rounds of selection, an NTD was obtained with specificity for a 5' G. Numerous selections were performed in attempts to obtain variants that recognized either 5' A or 5' C. The G230-K234 hairpin was inverted, the K230-G234/ins232 hairpin extended, modification of the K265-G268 N₀ hairpin attempted, and random mutagenesis libraries evaluated. None of these strategies yielded NTDs with affinity for target sequences with 5' A or 5' C, although we did identify an NTD, NT-βN, with a deletion that recognized substrates with both 5' A and 5' C residues with acceptable affinity. The strong selection preference exhibited by the NTDs NT-T and NT-G and the importance of W232 in NT-T and R232 in NT-G are likely due to specific interactions of these amino acids with the 5' terminal residue of the DNA recognition sequence. It was recently reported that the *Ralstonia solanacearum* TALE stringently requires a 5' G, and a sequence alignment with NT-G shows what appears to be a comparable N-1 hairpin containing an arginine at the position analogous to 232 in NT-G (Figure 25). Owing to the high structural homology between the NTDs Brg11 and NT-T, it may be possible to modify the preference of the *Ralstonia* TALE NTD to thymine by a simple arginine to tryptophan mutation or to eliminate specificity by grafting NT-αN or NT-βN domains into this related protein. It is also interesting to note that arginine-guanine interactions are common in evolved zinc finger domains.

[0165] The variant NTDs selected were successfully imported into TALE-TFs, MBP-TALEs and TALENs and generally conferred the activity and specificity expected based on data from the recombinase evolution system. TALE-TFs with optimized NTDs enhanced TALE activation between 400- and 1500-fold relative to the activity of NT-T against AvrXa7 promoter sites with non-T 5' residues. When incorporated into TALENs, our NTD with non-T selectivity enhanced activity 2-9-fold relative to that of the NT-T domain on substrates with 5' A, C or G. The increases in TALEN gene editing generally correlated with increases in activity observed in TALE-R and TALE-TF constructs. The specificity and high activity of NT-G was maintained, as evidenced by the lower activity in assays with TALEN pairs A1/A2, C1/C2, and T1/T2, and the generally high activity of NT-αN and NT-βN was also imparted into the TALEN Δ152/+63 architecture.

[0166] It was recently reported that alternatively truncated TALEs with synthetic TALE RVD domains do not require a 5' T in the DNA substrate. The reported Δ143, +47 truncation was constructed as a Goldy TALE-TF and substantially lower activity on the AvrXa7 substrate was observed than for the Δ127, +95 truncation, which has been most commonly used by others and which is the truncation set used in our study (Figure 29). Thus, the difference in reported outcomes could be due to the truncated architectures used.

[0167] In summary, the importance of a 5' thymidine in the DNA substrate for binding and activity of designed TALEs was determined in the context of TALE-R, TALE-TF, MBP-TALEs and TALEN chimeras. Targeted mutagenesis and TALE-R selection were applied to engineer TALE NTDs that recognize bases other than thymine as the 5' most base of the substrate DNA. The engineered TALE domains developed here demonstrated modularity and were highly active in TALE-TF and TALEN architectures. These novel NTDs expand by ~15-fold the number of sites that can be targeted by current TALE-Rs, which have strict geometric requirements on their binding sites and which are highly sensitive to the identity of the N₀ base. Furthermore, they now allow for the precise placement of TALE DBDs and TALE-TFs at any DNA sequence to facilitate gene regulation, displacement of endogenous DNA-binding proteins and synthetic biology applications

where precise binding might be key. Although TALENs based on the native NTD show varying degrees of tolerance of N₀ base substitutions, the data indicate that the novel NTDs reported here also facilitate higher efficiency gene editing with any N₀ base as compared with natural NTD-based TALENs.

EXAMPLE 5

CHIMERIC ZINC FINGER RECOMBINASES

[0168] The following materials and method were utilized.

[0169] The split gene reassembly vector (pBLA) was derived from pBluescriptII SK (-) (Stratagene) and modified to contain a chloramphenicol resistance gene and an interrupted TEM-1 p lactamase gene under the control of a *lac* promoter. ZFR target sites were introduced as previously described. Briefly, GFPuv (Clontech) was PCR amplified with the primers GFP- ZFR-XbaI-Fwd and GFP-ZFR-HindIII-Rev and cloned into the SpeI and HindIII restriction sites of pBLA to generate pBLA-ZFR substrates. All primer sequences are provided in Table 8.

Table 8. Primer Sequences.

Primer	SEQ ID NO:	Sequence
GFP-ZFR-20G-XbaI-Fwd	264	TTAATTAAGAGTCTAGAGGAGGCGTGTCCAAAACCATGGTTTACAGCAC GCCTCCAGATCTAGGAGGAATTTAAAATGAG
GFP-ZFR-20G-HindIII-Rev	265	ACTGACCTAGAGAAGCTTGGAGGCGTGCTGTAAACCATGGTTTTGGACA CGCCTCCCTGCAGTTAATTGTACAGTTCATC
SV40-ZFR-1-BglII-Fwd	266	TTAATTAAGAGAGATCTGCTGATGCAGATACAGAAACCAAGGTTTCTT ACTTGCTGCTGCGCGATCTGCATCTCAATTAGTCAGC
CMV-PstI-ZFR-1 Rev	267	CACCACCACGGATCCGCAGCAGCAAGTAAGAAAACCTTGGTTTCTGTAT CTGCATCAGCAATTTCGATAAGCCAGTAAGCAG
5' Gin-HBS-Koz	268	CACCACCACGCGCGCAAGCTTAGATCTGGCCCAGGCGGCCACCATGCT GATTGGCTATGTAAGGG
3' Gin-AgeI-Rev	269	CACCACCACACCGGTTCCCGATTTAGGTGGGCGAC

Primer	SEQ ID NO:	Sequence
ZFR-Target-1-Fwd	270	GTTCTGCTGCCAGGATCCACTAG
ZFR-Target-1-Rev	271	GCATGTGTCCAGATGCATAGG
ZFR-Target-2-Fwd	272	CACCTTCTCCCAGGATAAGG
ZFR-Target-2-Rev	273	GTTGGCCTGTATTCCTCTGG
ZFR-Target-3-Fwd	274	AATGAAGTTCCCTTGGCACTTC
ZFR-Target-3-Rev	275	CTGAAGGGTTTTAAGTGCAGAAG
CMV-Mid Prim-1	276	TGACGTCAATGACGGTAAATGG

ZFR targets are underlined.

[0170] To generate luciferase reporter plasmids, the SV40 promoter was PCR amplified from pGL3- Prm (Promega) with the primers SV40-ZFR-BglIII-Fwd and SV40-ZFR-HindIII-Rev. PCR products were digested with BglIII and HindIII and ligated into the same restriction sites of pGL3- Prm to generate pGL3-ZFR-1, 2, 3...18. The pBPS-ZFR donor plasmid was constructed as previously described with the following exception: the ZFR-1, 2 and 3 recombination sites were encoded by primers 3' CMV-PstI-ZFR-1, 2 or 3-Rev. Correct construction of each plasmid was verified by sequence analysis.

Recombination assays.

[0171] ZFRs were assembled by PCR as previously described. PCR products were digested with *SacI* and *XbaI* and ligated into the same restriction sites of pBLA. Ligations were transformed by electroporation into *Escherichia coli* T0P10F' (Invitrogen). After 1 hr recovery in SOC medium, cells were incubated with 5 mL SB medium with 30 $\mu\text{g mL}^{-1}$ chloramphenicol and cultured at 37°C. At 16 hr, cells were harvested; plasmid DNA was isolated by Mini-prep (Invitrogen) and 200 ng pBLA was used to transform *E. coli* T0P10F'. After 1 hr recovery in SOC, cells were plated on solid LB media with 30 $\mu\text{g mL}^{-1}$ chloramphenicol or 30 $\mu\text{g mL}^{-1}$ chloramphenicol and 100 $\mu\text{g mL}^{-1}$ carbenicillin, an ampicillin analogue. Recombination was determined as the number of colonies on LB media containing chloramphenicol and carbenicillin divided by the number of colonies on LB media containing chloramphenicol. Colony number was determined by automated counting using the GelDoc XR Imaging System (Bio-Rad).

Selections.

[0172] The ZFR library was constructed by overlap extension PCR as previously described. Mutations were introduced at positions 120, 123, 127, 136 and 137 with the degenerate codon NNK (N: A, T, C or G and K: G or T), which encodes all 20 amino acids. PCR products were digested with SacI and XbaI and ligated into the same restriction sites of pBLA. Ligations were ethanol precipitated and used to transform *E. coli* T0P10F'. Library size was routinely determined to be $\sim 5 \times 10^7$. After 1 hr recovery in SOC medium, cells were incubated in 100 mL SB medium with 30 $\mu\text{g mL}^{-1}$ chloramphenicol at 37°C. At 16 hr, 30 mL of cells were harvested; plasmid DNA was isolated by Mini-prep and 3 μg plasmid DNA was used to transform *E. coli* T0P10F'. After 1 hr recovery in SOC, cells were incubated with 100 mL SB medium with 30 $\mu\text{g mL}^{-1}$ chloramphenicol and 100 $\mu\text{g mL}^{-1}$ carbenicillin at 37°C. At 16 hr, cells were harvested and plasmid DNA was isolated by Maxi-prep (Invitrogen). Enriched ZFRs were isolated by SacI and XbaI digestion and ligated into fresh pBLA for further selection. After 4 rounds of selection, sequence analysis was performed on individual carbenicillin-resistant clones. Recombination assays were performed as described above.

ZFR construction.

[0173] Recombinase catalytic domains were PCR amplified from their respective pBLA selection vector with the primers 5' Gin-HBS-Koz and 3' Gin-Agel-Rev. PCR products were digested with HindIII and AgeI and ligated into the same restriction sites of pBH to generate the SuperZiF-compatible subcloning plasmids: pBH-Gin-a, P, y, 5, s or Z. Zinc-fingers were assembled by SuperZiF and ligated into the AgeI and SpeI restriction sites of pBH-Gin-a, P, y, 5, s or Z to generate pBH-ZFR-L/R-1, 2, 3.18 (L: left ZFR; R: right ZFR). ZFR genes were released from pBH by SfiI digestion and ligated into pcDNA 3.1 (Invitrogen) to generate pcDNA-ZFR-L/R-1, 2, 3.18. Correct construction of each ZFR was verified by sequence analysis (Table 9).

Table 9. Catalytic domain substitutions and intended DNA targets.

Positions						
Catalytic domain	Target	120	123	127	136	137
A	CC ^a	Ile	Thr	Leu	Ile	Gly
B	GC	Ile	Thr	Leu	Arg	Phe
Γ	GT	Leu	Val	Ile	Arg	Trp
Δ	CA	Ile	Val	Leu	Arg	Phe
ε ^b	AC	Leu	Pro	His	Arg	Phe
ζ ^c	TT	Ile	Thr	Arg	Ile	Phe

^aIndicates wild-type DNA target.

^bThe ε catalytic domain also contains the substitutions E117L and L118S.

^cThe ζ catalytic domain also contains the substitutions M124S, R131I and P141R.

Luciferase assays.

[0174] Human embryonic kidney (HEK) 293 and 293T cells (ATCC) were maintained in DMEM containing 10% (vol/vol) FBS and 1% (vol/vol) Antibiotic-Antimycotic (Anti-Anti; Gibco). HEK293 cells were seeded onto 96-well plates at a density of 4×10^4 cells per well and established in a humidified 5% CO₂ atmosphere at 37°C. At 24 hr after seeding, cells were transfected with 150 ng pcDNA-ZFR-L 1-18, 150 ng pcDNA-ZFR-R 1-18, 2.5 ng pGL3-ZFR-1, 2, 3. or 18 and 1 ng pRL-CMV using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. At 48 hr after transfection, cells were lysed with Passive Lysis Buffer (Promega) and luciferase expression was determined with the Dual-Luciferase Reporter Assay System (Promega) using a Veritas Microplate Luminometer (Turner Biosystems).

Integration assays.

[0175] HEK293 cells were seeded onto 6-well plates at a density of 5×10^5 cells per well and maintained in serum-containing media in a humidified 5% CO₂ atmosphere at 37°C. At 24 hr after seeding, cells were transfected with 1 μ g pcDNA-ZFR-L-1, 2 or 3 and 1 μ g pcDNA-ZFR- R-1, 2 or 3 and 200 ng pBPS-ZFR-1, 2 or 3 using Lipofectamine 2000 according to the manufacturer's instructions. At 48 hr after transfection, cells were split onto 6-well plates at a density of 5×10^4 cells per well and maintained in serum-containing media with 2 μ g mL⁻¹ puromycin. Cells were harvested upon reaching 100% confluence and genomic DNA was isolated with the Quick Extract DNA Extraction Solution (Epicentre). ZFR targets were PCR amplified with the following primer combinations: ZFR-Target-1, 2 or 3-Fwd and ZFR-Target-1, 2 or 3-Rev (Unmodified target); ZFR-Target-1, 2 or 3-Fwd and CMV-Mid-Prim-1 (Forward integration); and CMV-Mid-Prim-1 and ZFR-Target-1, 2 or 3-Rev (Reverse integration) using the Expand High Fidelity Taq System (Roche). For clonal analysis, at 2 days post-transfection 1×10^5 cells were split onto a 100 mm dish and maintained in serum-containing media with 2 μ g mL⁻¹ puromycin. Individual colonies were isolated with 10 mm X 10 mm open-ended cloning cylinders with sterile silicone grease (Millipore) and expanded in culture. Cells were harvested upon reaching 100% confluence and genomic DNA was isolated and used as template for PCR, as described above. For colony counting assays, at 2 days post-transfection cells were split into 6-well plates at a density of 1×10^4 cells per well and maintained in serum-containing media with or without 2 μ g mL⁻¹ puromycin. At 16 days, cells were stained with a 0.2% crystal violet solution and integration efficiency was determined by counting the number of colonies formed in puromycin-containing media divided by the number of colonies formed in the absence of puromycin. Colony number was determined by automated counting using the GelDoc XR Imaging System (Bio-Rad).

Results.

Specificity profile of the Gin recombinase.

[0176] In order to re-engineer serine recombinase catalytic specificity, a detailed understanding was developed of the factors underlying substrate recognition by this family of enzymes. To accomplish this, the ability of an activated mutant of the catalytic domain of the DNA invertase Gin to recombine a comprehensive set of symmetrically substituted target sites was evaluated. The Gin catalytic domain recombines a pseudo-symmetric 20-bp core that consists of two 10-bp half-site regions. This collection of recombination sites therefore contained each possible single-base substitution at positions 10, 9, 8, 7, 6, 5, and 4 and each possible two-base combination at positions 3 and 2 and in the dinucleotide core. Recombination was determined by split gene reassembly, a previously described method that links recombinase activity to antibiotic resistance.

[0177] In general, it was found that Gin tolerates (i) 12 of the 16 possible two-base combinations at the dinucleotide core (AA, AT, AC, AG, TA, TT, TC, TG, CA, CT, GA, GT); (ii) 4 of the 16 possible two-base combinations at positions 3 and 2 (CC, CG, GG and TG); (iii) a single A to T substitution at positions 6, 5, or 4; and (iv) all 12 possible single-base substitutions at positions 10, 9, 8, and 7 (Figure 31A-D). Further, it was found that Gin could recombine a target site library containing at least 10^6 (of a possible 4.29×10^9) unique base combinations at positions 10, 9, 8, and 7 (Figure 31D).

[0178] These findings are consistent with observations made from crystal structures of the yS resolvase, which indicate that (i) the interactions made by the recombinase dimer across the dinucleotide core are asymmetric and predominately non-specific; (ii) the interactions between an evolutionarily conserved Gly-Arg motif in the recombinase arm region and the DNA minor groove imposes a requirement for adenine or thymine at positions 6, 5, and 4; and (iii) there are no sequence-specific interactions between the arm region and the minor groove at positions 10, 9, 8, or 7 (Figure 31E). These results are also consistent with studies that focused on determining the DNA-binding properties of the closely related Hin recombinase.

Re-engineering Gin recombinase catalytic specificity.

[0179] Based on the finding that Gin tolerates conservative substitutions at positions 3 and 2 (i.e., CC, CG, GG, and TG), whether Gin catalytic specificity could be re-engineered to specifically recognize core sequences containing each of the 12 base combinations not tolerated by the native enzyme (Figure 32A) was investigated. In order to identify the specific amino acid residues involved in DNA recognition by Gin, the crystal structures of two related serine recombinases, the y6 resolvase and Sin recombinase, in complex with their respective DNA targets were examined. Based on these models, five residues were identified that contact DNA at positions 3 and 2: Leu 123, Thr 126, Arg 130, Val 139, and Phe 140 (numbered according to the y5 resolvase) (Figure 32B). Random mutagenesis was performed on the equivalent residues in the Gin catalytic domain (Ile 120, Thr 123, Leu 127, Ile 136, and Gly 137) by overlap extension PCR and

constructed a library of ZFR mutants by fusing these catalytic domain variants to an unmodified copy of the 'H1' ZFP. The theoretical size of this library was 3.3×10^7 variants.

[0180] The ZFR library was cloned into substrate plasmids containing one of the five base combinations not tolerated by the native enzyme (GC, GT, CA, AC, or TT) and enriched for active ZFRs by split gene reassembly (Figure 32C). After 4 rounds of selection, it was found that the activity of each ZFR population increased >1,000-fold on DNA targets containing GC, GT, CA, and TT substitutions and >100-fold on a DNA target containing AC substitutions (Figure 32D).

[0181] Individual recombinase variants were sequenced from each population and found that a high level of amino acid diversity was present at positions 120, 123, and 127 and that >80% of selected clones contained Arg at position 136 and Trp or Phe at position 137 (Figure 36). These results suggest that positions 136 and 137 play critical roles in the recognition of unnatural core sequences. The ability of each selected enzyme to recombine its target DNA was evaluated and it was found that nearly all recombinases showed activity (>10% recombination) and displayed a >1,000-fold shift in specificity toward their intended core sequence (Figure 37). As with the parental Gin, it was found that several recombinases tolerated conservative substitutions at positions 3 and 2 (i.e., cross-reactivity against GT and CT or AC and AG), indicating that a single re-engineered catalytic domain could be used to target multiple core sites (Figure 37).

[0182] In order to further investigate recombinase specificity, the recombination profiles were determined of five Gin variants (hereafter designated Gin p, y, 6, e and Z) shown to recognize nine of the 12 possible two-base combinations not tolerated by the parental enzyme (GC, TC, GT, CT, GA, CA, AG, AC, and TT) (Table 1). Gin p, 6, and e recombined their intended core sequences with activity and specificity comparable to that of the parental enzyme (hereafter referred to as Gin a) and that Gin y and Z were able to recombine their intended core sequences with specificity exceeding that of Gin a (Figure 32E). Each recombinase displayed >1,000-fold preference for adenine or thymine at positions 6, 5, and 4 and showed no base preference at positions 10, 9, 8, and 7 (Figure 38). These results indicate that mutagenesis of the DNA-binding arm did not compromise recombinase specificity. It was not possible to select for Gin variants capable of tolerating AA, AT, or TA substitutions at positions 3 and 2. One possibility for this result is that DNA targets containing >4 consecutive A-T bps might exhibit bent DNA conformations that interfere with recombinase binding and/or catalysis.

Engineering ZFRs to recombine user-defined sequences

[0183] Whether ZFRs composed of the re-engineered catalytic domains could recombine pre-determined sequences was investigated. To test this possibility, the human genome (GRCh37 primary reference assembly) was searched for potential ZFR target sites using a 44-bp consensus recombination site predicted to occur approximately once every 400,000 bp of random DNA (Fig. 4A). This ZFR consensus target site, which was derived from the core sequence profiles of the selected Gin variants, includes approximately 7×10^8 (of a possible 1.0955×10^{12}) unique 20-bp core combinations predicted to be tolerated by the 21 possible

catalytic domain combinations and a conservative selection of modular zinc finger domains that excludes 5'-CNN-3' and 5'-TNN-3' triplets within each ZFBS. Using ZFP specificity as the primary determinant for selection, 18 possible ZFR target sites across 8 human chromosomes (Chr. 1, 2, 4, 6, 7, 11, 13 and X) at non-protein coding loci were identified. On average, each 20- bp core showed ~46% sequence identity to the core sequence recognized by the native Gin catalytic domain (Figure 33B). Each corresponding ZFR was constructed by modular assembly (see Materials and Methods).

[0184] To determine whether each ZFR pair could recombine its intended DNA target, a transient reporter assay was developed that correlates ZFR-mediated recombination to reduced luciferase expression (Figures 33A and 39). To accomplish this, ZFR target sites were introduced upstream and downstream an SV40 promoter that drives expression of a luciferase reporter gene. Human embryonic kidney (HEK) 293T cells were co-transfected with expression vectors for each ZFR pair and its corresponding reporter plasmid. Luciferase expression was measured 48 hr after transfection. Of the 18 ZFR pairs analyzed, 38% (7 of 18) reduced luciferase expression by >75-fold and 22% (4 of 18) decreased luciferase expression by >140-fold (Figure 33B). In comparison, GinC4, a positive ZFR control designed to target the core sequence recognized by the native Gin catalytic domain, reduced luciferase expression by 107 fold. Overall, it was found that 50% (9 of 18) of the evaluated ZFR pairs decreased luciferase expression by at least 20-fold. Importantly, virtually every catalytic domain that displayed significant activity in bacterial cells (>20% recombination) was successfully used to recombine at least one naturally occurring sequence in mammalian cells.

[0185] In order to evaluate ZFR specificity, separately HEK293T cells were co-transfected with expression plasmids for the nine most active ZFRs with each non-cognate reporter plasmid. Each ZFR pair demonstrated high specificity for its intended DNA target and 77% (7 of 9) of the evaluated ZFRs showed an overall recombination specificity nearly identical to that of the positive control GinC4 (Fig. 4C). To establish that reduced luciferase expression is the product of the intended ZFR heterodimer and not the byproduct of recombination-competent ZFR homodimers, the contribution of each ZFR monomer to recombination was measured. Co- transfection of the ZFR 1 'left' monomer with its corresponding reporter plasmid led to a modest reduction in luciferase expression (total contribution to recombination: -22%), but the vast majority of individual ZFR monomers (16 of 18) did not significantly contribute to recombination (<10% recombination), and many (7 of 18) showed no activity (Figure 39). Taken together, these studies indicate that ZFRs can be engineered to recombine user-defined sequences with high specificity.

Engineered ZFRs mediate targeted integration into the human genome.

[0186] Whether ZFRs could integrate DNA into endogenous loci in human cells was evaluated next. To accomplish this, HEK293 cells were co-transfected with ZFR expression vectors and a corresponding DNA donor plasmid that contained a specific ZFR target site and a puromycin-resistance gene under the control of an SV40 promoter. For this analysis, ZFR pairs 1, 2, and 3, were used which were designed to target non-protein coding loci on human chromosomes 4, X, and 4, respectively (Figure 34A). At 2 days post-transfection, cells were incubated with

puromycin- containing media and measured integration efficiency by determining the number of puromycin- resistant (puro^R) colonies. It was found that (i) co-transfection of the donor plasmid and the corresponding ZFR pair led to a >12-fold increase in puro^R colonies in comparison to transfection with donor plasmid only and that (ii) co-transfection with both ZFRs led to a 6-to 9fold increase in puro^R colonies in comparison to transfection with individual ZFR monomers (Figure 34B). In order to evaluate whether ZFR pairs correctly targeted integration, genomic DNA was isolated from puro^R populations and amplified each targeted locus by PCR. The PCR products corresponding to integration in the forward and/or reverse orientations were observed at each locus targeted by these ZFR pairs (Figure 34C). Next, to determine the overall specificity of ZFR-mediated integration, genomic DNA was isolated from clonal cell populations and evaluated plasmid insertion by PCR. This analysis revealed targeting efficiencies of 8.3% (1 of 12 clones), 14.2% (5 of 35 clones), and 9.1% (1 of 11 clones) for ZFR pairs 1, 2, and 3, respectively (Fig. S6). Sequence analysis of each PCR product confirmed ZFR-mediated integration (Figure 34D). Taken together, these results indicate that ZFRs can be designed to accurately integrate DNA into endogenous loci.

[0187] Finally, it is noted that the ZFR-1 'left' monomer was found to target integration into the ZFR-1 locus (Figure 34C). This result, which is consistent with the luciferase reporter studies described above (Figure 39) indicates that recombination-competent ZFR homodimers have the capacity to mediate off-target integration. Future development of an optimized heterodimeric ZFR architecture and a comprehensive evaluation of off-target integration should lead to the design of ZFRs that demonstrate greater targeting efficiency.

[0188] It is herein shown that ZFRs can be designed to recombine user-defined sequences with high specificity and that ZFRs can integrate DNA into pre-determined endogenous loci in human cells. By combining substrate specificity analysis and directed evolution, virtually all sequence requirements imposed by the ZFR catalytic domain were eliminated. Using the archive of 45 pre-selected zinc-finger modules, it is estimated that ZFRs can be designed to recognize $>1 \times 10^{22}$ unique 44-bp DNA sequences, which corresponds to approximately one potential ZFR target site for every 4,000 bp of random sequence. Construction of customized zinc-finger domains by selection would further extend targeting. The re-engineered catalytic domains described herein will be compatible with recently described TAL effector recombinases. This work demonstrates the feasibility of generating ZFRs with custom specificity and illustrates the potential utility of ZFRs for a wide range of applications, including genome engineering, synthetic biology, and gene therapy.

SEQUENCE LISTING

[0189]

<110> THE SCRIPPS RESEARCH INSTITUTE
 BARBAS, Carlos F. III
 MERCER, Andrew

LAMB, Brian M.

GAJ, Thomas

<120> CHIMERIC POLYPEPTIDES HAVING TARGETED BINDING SPECIFICITY

<130> SCRIP2070-2WO

<150> US 61/818,364

<151> 2013-05-01

<150> US 61/753,763

<151> 2013-01-17

<150> US 61/696,689

<151> 2012-09-04

<160> 369

<170> PatentIn version 3.5

<210> 1

<211> 864

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 1

atggatccca ttcgttcgcg cacgccaagt cctgcccgcg agcttctgcc cggaccccaa	60
ccggataggg ttcagccgac tgcagatcgg gggggggctc cgctgtctgg cggccccctg	120
gatggcttgc ccgctcggcg gacgatgtcc cggaccgcgc tgccatctcc ccctgcgccc	180
tcgcctgcgt tctcggcggg cagcttcagc gatctgtccc gtcagttcga tccgtcgttt	240
cttgatacat cgcttcttga ttcgatgcct gccgtcggca cgccgcatac agcggctgcc	300
ccagcagagt gcgatgaggt gcaatcgggt ctgcgtgcag ccgatgacct gccacccacc	360
gtgcgtgtcg ctgtcactgc ggcgcggccg ccgcgcgccca agccggcccc gcgacggcgt	420
gcggcgcaac cctccgacgc ttcgccggcc gcgcaggtgg atctacgcac gctcggctac	480
agtcagcagc agcaagagaa gatcaaaccg aaggtgcgtt cgacagtggc gcagcaccac	540
gaggcactgg tggggccatgg gtttacacac gcgcacatcg ttgcgctcag ccaacacccg	600
gcagcgttag ggaccgttgc tgtcacgtat caggacataa tcagggcggt gccagaggcg	660
acacacgaag acatcgttgg cgtcggcaaa cagtgggtccg gcgcacgcgc tctggaggcc	720
ttgctcacgg aggcggggga gttgagaggt ccgccgttac agttggacac aggccaactt	780
ctcaagattg caaaacgtgg cggcgtgacc gcagtggagg cagtgcctgc atggcgcaat	840
gcactgacgg gtgccccctt gaac	864

<210> 2

<211> 288

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 2

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

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

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

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

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

Leu	Asp	Thr	Ser	Leu	Leu	Asp	Ser	Met	Pro	Ala	Val	Gly	Thr	Pro	His
				85				90						95	

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

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

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

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

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

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

Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala	Val
		195					200					205			

Thr	Tyr	Gln	Asp	Ile	Ile	Arg	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Asp
	210					215					220				

Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala
225					230					235					240

Leu	Leu	Thr	Glu	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp
				245					250					255	

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

Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn
275 280 285

<210> 3

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> MISC_FEATURE

<222> (4)..(4)

<223> Xaa is Y, S or R

<220>

<221> MISC_FEATURE

<222> (5)..(5)

<223> Xaa is R or G

<220>

<221> MISC_FEATURE

<222> (6)..(6)

<223> Xaa is R, H, A, N or T

<400> 3

Val Gly Lys Xaa Xaa Xaa Gly Ala Arg Ala Leu
1 5 10

<210> 4

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 4

Val Gly Lys Tyr Arg Gly Ala Arg Ala Leu
1 5 10

<210> 5

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 5

Val	Gly	Lys	Ser	Arg	Ser	Gly	Ala	Arg	Ala	Leu
1				5					10	

<210> 6

<211> 10

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 6

Val	Gly	Lys	Tyr	His	Gly	Ala	Arg	Ala	Leu
1				5					10

<210> 7

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 7

Val	Gly	Lys	Arg	Gly	Ala	Gly	Ala	Arg	Ala	Leu
1				5					10	

<210> 8

<211> 13

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> MISC_FEATURE

<222> (7)..(7)

<223> Xaa is Y, S or R

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> Xaa is W or G

<220>

<221> MISC_FEATURE

<222> (9)..(9)

<223> Xaa is R, H, A, N or T

<400> 8

Ile	Val	Asp	Ile	Ala	Lys	Xaa	Xaa	Xaa	Gly	Asp	Leu	Ala
1				5					10			

<210> 9

<211> 13

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 9

Ile	Val	Asp	Ile	Ala	Arg	Gln	Trp	Ser	Gly	Asp	Leu	Ala
1				5					10			

<210> 10

<211> 12

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 10

Ile	Val	Asp	Ile	Ala	Arg	Tyr	Arg	Gly	Asp	Leu	Ala
1				5					10		

<210> 11

<211> 13

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 11

Ile	Val	Asp	Ile	Ala	Arg	Ser	Arg	Ser	Gly	Asp	Leu	Ala
1				5					10			

<210> 12

<211> 12

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 12

Ile	Val	Asp	Ile	Ala	Arg	Tyr	His	Gly	Asp	Leu	Ala
1				5					10		

<210> 13

<211> 13
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 13
 Ile Val Asp Ile Ala Arg Arg Gly Ala Gly Asp Leu Ala
 1 5 10

<210> 14
 <211> 9
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<220>
 <221> misc_feature
 <222> (4)..(6)
 <223> Xaa can be any naturally occurring amino acid

<400> 14
 Val Gly Lys Xaa Xaa Xaa Gly Ala Arg
 1 5

<210> 15
 <211> 12
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<220>
 <221> misc_feature
 <222> (5)..(8)
 <223> Xaa can be any naturally occurring amino acid

<400> 15
 Val Asp Ile Ala Xaa Xaa Xaa Xaa Gly Asp Leu Ala
 1 5 10

<210> 16
 <211> 48
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 16

Asn	Ile	His	Gly	Asn	Ile	Asn	Ile	Asn	Ser	His	Asp	Asn	Asn	His	Asp
1				5					10					15	

His	Asp	His	Asp	Asn	Ser	Asn	Asn	His	Asp	His	Asp	Asn	Ser	Asn	Ser
			20					25					30		

Asn	Asn	Asn	Asn	Asn	Ile	Asn	Gly	Asn	Asn	Asn	Ile	Asn	Asn	Ser	Asn
		35					40					45			

<210> 17

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 17

ataaaccccc tccaaccagg tgctaa 26

<210> 18

<211> 52

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 18

Asn	Ile	Asn	Gly	Asn	Ile	Asn	Ile	Asn	Ile	His	Asp	His	Asp	His	Asp
1				5					10					15	

His	Asp	His	Asp	Asn	Gly	His	Asp	His	Asp	Asn	Ile	Asn	Ile	His	Asp
			20					25					30		

His	Asp	Asn	Ile	Asn	Asn	Asn	Asn	Asn	Gly	Asn	Asn	His	Asp	Asn	Gly
		35					40					45			

Asn	Ile	Asn	Ile
			50

<210> 19

<211> 1446

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 19

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

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

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

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

```

          945          950          955          960
Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val
945          950          955          960

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp
          965          970          975

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu
          980          985          990

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
          995          1000          1005

Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
1010          1015          1020

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
1025          1030          1035

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
1040          1045          1050

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
1055          1060          1065

Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala Ile
1070          1075          1080

Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
1085          1090          1095

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn Gln Val
1100          1105          1110

Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr
1115          1120          1125

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
1130          1135          1140

Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala
1145          1150          1155

Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu
1160          1165          1170

Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
1175          1180          1185

Gly Arg Pro Ala Leu Asp Ala Val Lys Lys Gly Leu Pro His Ala
1190          1195          1200

Pro Glu Leu Ile Arg Arg Ile Asn Arg Arg Ile Pro Glu Arg Thr
1205          1210          1215

Ser His Arg Val Pro Asp Leu Ala His Val Val Arg Val Leu Gly
1220          1225          1230

```

<211> 4341

<213> Artificial Sequence

<223> Synthetic construct

atggatccca ttogttcgcg cacgccaaagt cctgcccgcg agcttctgcc cggaccccaa	60
ccggataggg ttcagccgac tgcagatcgg gggggggctc cgcttctgg cggccccctg	120
gatggcttgc ccgctcggcg gacgatgtcc cggacccggc tgccatctcc cctgcgccc	180
tcgcctgcgt tctcggcggg cagcttcagc gatctgctcc gtcagttcga tccgtcgtt	240
cttgatacat cgcttcttga ttcgatgcct gccgtcggca cgcgcatac agcggctgcc	300
ccagcagagt gcgatgaggt gcaatcgggt ctgcgtgcag ccgatgacct gccaccacc	360
gtgcgtgtcg ctgtcactgc ggcgcggccg ccgcgcgcca agccggcccc gcgacggcgt	420
gcggcgcaac cctccgacgc ttcgccggcc gcgcaggtgg atctacgcac gctcggctac	480
agtcagcagc agcaagagaa gatcaaaccg aaggtgcgtt cgacagtggc gcagcaccac	540
gaggcactgg tgggccatgg gtttacacac gcgcacatcg ttgcgtcag ccaacacccg	600
gcagcgttag ggaccgttgc tgtcacgtat caggacataa tcagggcgtt gccagaggcg	660
acacacgaag acatcgttgg cgtcggcaaa cagtgggccg gcgcacgcgc tctggaggcc	720
ttgctcacgg aggcggggga gttgagaggt ccgcgttac agttggacac aggccaaactt	780
ctcaagattg caaaacgtgg cggcgtgacc gcagtggagg cagtgcattc atggcgcaat	840
gcactgacgg gtgccccct gaacctgacc ccggaccaag tgggtggccat cgcagcaat	900
attggcggca agcaggcgtt ggagacggta cagcggctgt tgccggtgct gtgccaggac	960
catggcctga ccccggaaca ggtcgtggcc atcgccagcc atggcgggcg caagcaggcg	1020
ctggagacgg tgcagcggct gttgccggtg ctgtgccagg accatggcct gaccccgga	1080
caggtggtgg ccatcgccag caatattggc ggcaagcagg cgctagagac ggtgcagcgg	1140
ctgttgccgg tgctgtgcca ggcccatggc ctgaccccg accaggtcgt ggccatcgcc	1200
agcaatattg gcggcaagca ggcgctggag acggtgcagc ggctgttgcc ggtgctgtgc	1260
caggaccatg gcctgacccc ggcccagggt gtggccatcg ccagcaatag tggcggcaag	1320
caggcgttgg agacggtgca gcggctgttg ccggtgctgt gccaggacca tggcctgacc	1380
ccggaccaag tcgtggccat cgcagccac gatggcggca agcaggcgtt ggagacgctg	1440
cagcggctgt tgccggtgct gtgccaggac catggcctga ccccggaaca ggtcgtggcc	1500
atcgccaaca ataacggcgg caagcaggcg ctggagacgc tgcagcggct gttgccggtg	1560
ctgtgccagg accatggcct gaccccgga caagtgggtg ccatcgccag ccacgatggc	1620
ggcaagcagg cgctggagac ggtgcagcgg ctgttgccgg tgctgtgcca ggaccatggc	1680
ctgaccccg accaggtggt ggccatcgcc agccacgatg gcggcaagca ggcgctggag	1740
acggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gcctgacccc ggcccaagtg	1800
gtggccatcg ccagccacga tggcggcaag caggcgttgg agacggtgca gcggctgttg	1860
ccggtgctgt gccaggacca tggcctgacc ccggaccagg tgggtggccat cgcagcaat	1920
agcggcggca agcaggcgtt ggagacggta cagcggctgt tgccggtgct gtgccaggac	1980
catggactga ccccggaaca ggtcgtggcc atcgccagca atggcggcaa gcaggcgtg	2040
gagacggtac agcggctgtt gccggtgctg tgccaggacc atggcctgac cccggaccag	2100
gtcgtggcca tcgccagcaa tggcggcaag caggcgttgg agacggtgca gcggctgttg	2160
ccggtacagc ggctgttgcc ggtgctgtgc caggaccatg gcctgaccca ggaccagggtg	2220
gtggccatcg ccagccacga tggcggcaag caggcgttgg agacggtgca gcggctgttg	2280

ccggtgctgt gccaggacca tggcctgacc ccggaccaag tgggtggccat cggcagccac	2340
gatggcggca aacaggcgct ggagacgggtg cagcggctgt tgccgggtgct gtgccaggac	2400
catggcctga ccccggaacca ggtggtggcc atcgccagca atagtggcgg caagcaggcg	2460
ctggagacgg tgcagcggct gttgccgggtg ctgtgccagg accatggcct gaccccgga	2520
caagtgggtg ccatcgccag caatagtggc ggcaagcagg cgctggagac ggtgcagcgg	2580
ctgttgccgg tgctgtgcca ggaccatggc ctgaccccg accaggtggt ggccatcgcc	2640
agcaataacg gcggcaagca ggcgctggag acggtgcagc ggctgttgcc ggtgctgtgc	2700
caggaccatg gcctgacccc ggaccaggtc gtggccatcg ccaacaataa cggcggcaag	2760
caggcgctgg agacgggtgca ggcgctgttg ccggtgctgt gccaggacca tggcctgacc	2820
ccggcgccagg tgggtggccat cggcagcaat attggcggca agcaggcgct ggagacgggtg	2880
cagcggctgt tgccgggtgct gtgccaggac catggcctga ccctggacca ggtggtggcc	2940
attgccagca atggcggcag caaacaggcg ctagagacgg tgcagcggct gttgccgggtg	3000
ctgtgccagg accatggcct gaccccgga caagtgggtg ccatcgccaa caataacggc	3060
ggcaagcagg cgctggagac ggtgcagcgg ctgttgccgg tgctgtgcca ggaccatggc	3120
ctgaccccg accaggtcgt ggccatcgcc agcaatattg gcggcaagca ggcgctggag	3180
acggtgcagc ggctgttgcc ggtgctgtgc caggaccatg gcctgaccct ggaccaggtg	3240
gtggccatcg ccagcaatgg cggcaagcag gcgctggaga cgggtgcagc gctgttgccg	3300
gtgctgtgcc aggaccatgg cctgacccc aaccaggtgg tggccatcgc cagcaatagt	3360
ggcggcaagc aggcgctgga gacgggtgag cggctgttgc cgggtgctgt ccaggaccat	3420
ggcctgaccc cgaaccaggt ggtggccatc gccagcaatg gcggcaagca ggcgctggag	3480
agcattgttg ccagttatc tcgccctgat ccggcggttg ccgcgttgac caacgaccac	3540
ctcgtcgctt tggcctgcct cggcggacgt cctgccctgg atgcagtga aaagggttg	3600
ccgcacgcgc cgggaattgat cagaagaatc aatcgccgca tccccgaac cacgtcccat	3660
cgcgttcccc acctcgcgca cgtggttcgc gtgcttggtt ttttcagag ccactccac	3720
ccagcgcaag cattcgatga cggcatgacg cagttcgaga tgagcaggca cggcttggtg	3780
cagctctttc gcagagtggg cgtcaccgaa ttcgaagccc gctacggaac gctccccca	3840
gcctcgcagc gttgggaccg taccctccag gcatcaggga tgaaaagggc caaacgctcc	3900
cctacttcag ctcaaaccac ggatcaggcg tctttgcatg cattcgccga ttcgctggag	3960
cgtgaccttg atgcgcccag cccaatgcac gagggagatc agacgcgggc aagcagccgt	4020
aaacgggtccc gatcggatcg tgctgtcacc ggccctcca cacagcaatc tttcgaggtg	4080
cgcgttcccc aacagcaaga tgcgctgcat ttgccctca gctggagggt aaaacgccc	4140
cgtaccagga tcgggggagg cctcccgat cctggtacgc ccatcgctgc cgacctggca	4200
gcgtccagca ccgtgatgtg ggaacaagat gcggccccct tcgcaggggc agcgatgat	4260
ttcccgcat tcaacgaaga ggagctcgca tggttgatgg agctattgcc tcagtcaggc	4320
tcagtcggag ggacgatctg a	4341

<210> 21

<211> 1263

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 21

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly
130 135 140

Ser Gly Ser Pro Arg Gln Phe Asp Pro Ser Leu Leu Asp Thr Ser Leu
145 150 155 160

Leu Asp Ser Met Pro Ala Val Gly Thr Pro His Thr Ala Ala Ala Pro
165 170 175

Ala Glu Cys Asp Glu Val Gln Ser Gly Leu Arg Ala Ala Asp Asp Pro
180 185 190

Pro Pro Thr Val Arg Val Ala Val Thr Ala Ala Arg Pro Pro Arg Ala
195 200 205

Lys Pro Ala Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro
210 215 220

Ala Ala Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln
225 230 235 240

Glu Lys Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu
245 250 255

Ala Leu Val Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser
260 265 270

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

Ile	Arg	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	
290						295				300						
Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Glu	Ala	
305				310					315						320	
Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Leu	
				325				330						335		
Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Val	Glu	Ala	Val	His	Ala	
			340					345						350		
Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro	Leu	Asn	Leu	Thr	Pro	Asp	Gln	
		355					360					365				
Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	
	370					375					380					
Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	
385					390					395					400	
Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Gly	Gly	Gly	Lys	Gln	Ala	Leu	
				405					410						415	
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	
			420					425					430			
Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	
		435					440					445				
Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Ala	His	
	450					455					460					
Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	
465					470				475						480	
Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	
				485				490						495		
Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	
			500					505					510			
Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	
		515					520					525				
Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	
	530					535					540					
His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Leu	Gln	Arg	Leu	Leu	Pro	
545					550					555					560	
Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	
				565					570					575		
Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Leu	Gln	Arg	Leu	
			580					585					590			

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val
 595 600 605

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 610 615 620

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 625 630 635 640

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 645 650 655

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 660 665 670

Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu
 675 680 685

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
 690 695 700

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln
 705 710 715 720

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His
 725 730 735

Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys
 740 745 750

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 755 760 765

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly
 770 775 780

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Gln Arg Leu
 785 790 795 800

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Gln Asp Gln Val Val
 805 810 815

Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 820 825 830

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 835 840 845

Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr
 850 855 860

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 865 870 875 880

Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu
 885 890 895

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
 900 905 910
 Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln
 915 920 925
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His
 930 935 940
 Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly Gly
 945 950 955 960
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 965 970 975
 Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn
 980 985 990
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 995 1000 1005
 Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 1010 1015 1020
 Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 1025 1030 1035
 Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val
 1040 1045 1050
 Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu Thr
 1055 1060 1065
 Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 1070 1075 1080
 Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln
 1085 1090 1095
 Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 1100 1105 1110
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile
 1115 1120 1125
 Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 1130 1135 1140
 Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala Ile
 1145 1150 1155
 Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 1160 1165 1170
 Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn Gln Val
 1175 1180 1185
 Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr
 1190 1195 1200

1190		1195		1200
Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr				
1205		1210		1215
Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala				
1220		1225		1230
Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu				
1235		1240		1245
Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly				
1250		1255		1260

<210> 22

<211> 1250

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 22

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

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

Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 500 505 510

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 515 520 525

Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg
 530 535 540

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 545 550 555 560

Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Leu
 565 570 575

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 580 585 590

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu
 595 600 605

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 610 615 620

Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala
 625 630 635 640

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 645 650 655

Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 660 665 670

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 675 680 685

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly
 690 695 700

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 705 710 715 720

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 725 730 735

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 740 745 750

Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser
 755 760 765

Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 770 775 780

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Gln Asp
 785 790 795 800

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu

				805					810							815
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	
			820						825					830		
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	
		835					840					845				
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	
	850					855						860				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	
	865				870					875					880	
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	
				885					890						895	
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	
			900					905						910		
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	
		915					920						925			
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	
	930					935					940					
Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	
	945				950					955					960	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	
				965					970						975	
Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	
			980					985					990			
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val	Val	Ala	
		995					1000					1005				
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln		
	1010					1015						1020				
Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Leu	Asp		
	1025					1030						1035				
Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Ser	Lys	Gln	Ala	Leu		
	1040					1045						1050				
Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly		
	1055					1060						1065				
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly		
	1070					1075						1080				
Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys		
	1085					1090						1095				
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser		

1100	1105	1110
Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
1115	1120	1125
Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val		
1130	1135	1140
Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln		
1145	1150	1155
Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn		
1160	1165	1170
Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu		
1175	1180	1185
Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
1190	1195	1200
Leu Thr Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys		
1205	1210	1215
Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro		
1220	1225	1230
Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys		
1235	1240	1245
Leu Gly		
1250		

<210> 23

<211> 1216

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 23

Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp
1				5					10					15	
Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu
			20					25					30		
Asp	Lys	Leu	Ser	Gly	Thr	Arg	Thr	Asp	Arg	Pro	Gly	Leu	Lys	Arg	Ala
		35					40					45			
Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp
	50					55					60				
Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu
65					70					75					80

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

385		390		395		400
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Ala						
		405		410		415
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly						
		420		425		430
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys						
		435		440		445
Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser Asn						
		450		455		460
Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val						
		465		470		480
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala						
		485		490		495
Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg Leu Leu						
		500		505		510
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala						
		515		520		525
Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Leu Gln Arg						
		530		535		540
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val						
		545		550		560
Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val						
		565		570		575
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp						
		580		585		590
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu						
		595		600		605
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr						
		610		615		620
Pro Ala Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala						
		625		630		640
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly						
		645		650		655
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys						
		660		665		670
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp						
		675		680		685
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser 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 Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 725 730 735
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Gln Arg
 740 745 750
 Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Gln Asp Gln Val
 755 760 765
 Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val
 770 775 780
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 785 790 795 800
 Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu
 805 810 815
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 820 825 830
 Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala
 835 840 845
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 850 855 860
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys
 865 870 875 880
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 885 890 895
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly
 900 905 910
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 915 920 925
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn
 930 935 940
 Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 945 950 955 960
 Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala
 965 970 975
 Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 980 985 990
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala
 995 1000 1005

Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu Thr Val Gln
1010 1015 1020

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
1025 1030 1035

Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu
1040 1045 1050

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
1055 1060 1065

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly
1070 1075 1080

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
1085 1090 1095

Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala Ile Ala Ser
1100 1105 1110

Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
1115 1120 1125

Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn Gln Val Val Ala
1130 1135 1140

Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
1145 1150 1155

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn
1160 1165 1170

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu
1175 1180 1185

Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala
1190 1195 1200

Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
1205 1210 1215

<210> 24

<211> 1204

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 24

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

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

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 645 650 655
 Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 660 665 670
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 675 680 685
 Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 690 695 700
 Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 705 710 715 720
 Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 725 730 735
 Pro Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 740 745 750
 Gln Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala
 755 760 765
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 770 775 780
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 785 790 795 800
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 805 810 815
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ser Gly
 820 825 830
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 835 840 845
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 850 855 860
 Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 865 870 875 880
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 885 890 895
 Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 900 905 910
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 915 920 925
 Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 930 935 940
 Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala Gln Val

Leu Leu Phe Val Leu Cys Gln Asp His Gly Leu Thr Phe Ala Gln Val
 945 950 955 960

Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val
 965 970 975

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp
 980 985 990

Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala Leu Glu
 995 1000 1005

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
 1010 1015 1020

Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly Lys
 1025 1030 1035

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
 1040 1045 1050

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 1055 1060 1065

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
 1070 1075 1080

Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val Ala
 1085 1090 1095

Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 1100 1105 1110

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn Gln
 1115 1120 1125

Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu
 1130 1135 1140

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
 1145 1150 1155

Thr Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys Gln
 1160 1165 1170

Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
 1175 1180 1185

Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu
 1190 1195 1200

Gly

<210> 25

<211> 1190

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 25

Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp
1				5					10					15	

Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu
			20					25					30		

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

Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp
	50					55					60				

Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu
65					70					75					80

Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr
				85					90					95	

Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala
			100					105					110		

Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala
		115					120					125			

Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Arg	Lys	Ser	Gly
	130					135					140				

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

Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser
			165						170					175	

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

Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val
		195					200					205			

Ala	Val	Thr	Tyr	Gln	Asp	Ile	Ile	Arg	Ala	Leu	Pro	Glu	Ala	Thr	His
	210					215					220				

Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu
225					230					235					240

Glu	Ala	Leu	Leu	Thr	Glu	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu	Gln
				245					250					255	

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

Ala	Val	Glu	Ala	Val	His	Ala	Trp	Arg	Asn	Ala	Leu	Thr	Gly	Ala	Pro
		---					---					---			

275					280					285					
Leu	Asn	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly
290						295					300				
Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys
305					310					315					320
Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His
				325					330					335	
Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
			340					345					350		
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
		355					360					365			
Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
370						375					380				
Pro	Val	Leu	Cys	Gln	Ala	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
385				390						395					400
Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
				405				410						415	
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Ala	Gln	Val
			420					425					430		
Val	Ala	Ile	Ala	Ser	Asn	Ser	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
		435					440					445			
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
450						455					460				
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu
465						470					475				480
Thr	Leu	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
				485					490					495	
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Asn	Asn	Asn	Gly	Gly	Lys	Gln	Ala
			500					505					510		
Leu	Glu	Thr	Leu	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
		515					520					525			
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys
530						535					540				
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
545					550					555					560
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly
				565					570					575	
Gln	Leu	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 580 585 590

Gln Asp His Gly Leu Thr Pro Ala Gln Val Val Ala Ile Ala Ser His
 595 600 605

Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 610 615 620

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 625 630 635 640

Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 645 650 655

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 660 665 670

Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 675 680 685

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val
 690 695 700

Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 705 710 715 720

Leu Leu Pro Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 725 730 735

Leu Thr Gln Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 740 745 750

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 755 760 765

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly
 770 775 780

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 785 790 795 800

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 805 810 815

Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 820 825 830

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 835 840 845

Ser Asn Ser Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 850 855 860

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 865 870 875 880

Ile Ala Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 885 890 895

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 900 905 910
 Val Ala Ile Ala Asn Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val
 915 920 925
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Ala
 930 935 940
 Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
 945 950 955 960
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 965 970 975
 Leu Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Ser Lys Gln Ala
 980 985 990
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 995 1000 1005
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Asn Asn Asn Gly Gly
 1010 1015 1020
 Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 1025 1030 1035
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser
 1040 1045 1050
 Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 1055 1060 1065
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Leu Asp Gln Val Val
 1070 1075 1080
 Ala Ile Ala Ser Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 1085 1090 1095
 Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asn
 1100 1105 1110
 Gln Val Val Ala Ile Ala Ser Asn Ser Gly Gly Lys Gln Ala Leu
 1115 1120 1125
 Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 1130 1135 1140
 Leu Thr Pro Asn Gln Val Val Ala Ile Ala Ser Asn Gly Gly Lys
 1145 1150 1155
 Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro
 1160 1165 1170
 Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys
 1175 1180 1185

Leu Gly
1190

<210> 26

<211> 841

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 26

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly
130 135 140

Ser Gly Ser Pro Ala Leu Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg
145 150 155 160

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

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

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

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

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

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

545					550					555					560
Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val
				565					570					575	
Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala
			580					585					590		
Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu
		595					600					605			
Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala
	610					615					620				
Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg
625					630					635					640
Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp	Gln	Val
				645					650					655	
Val	Ala	Ile	Ala	Ser	Asn	Gly	Gly	Gly	Lys	Gln	Ala	Leu	Glu	Thr	Val
			660					665					670		
Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr	Pro	Asp
		675					680					685			
Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala	Leu	Glu
	690					695					700				
Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly	Leu	Thr
705					710					715					720
Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	His	Asp	Gly	Gly	Lys	Gln	Ala
					725									730	735
Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp	His	Gly
			740					745					750		
Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly	Gly	Lys
		755					760					765			
Gln	Ala	Leu	Glu	Thr	Val	Gln	Arg	Leu	Leu	Pro	Val	Leu	Cys	Gln	Asp
	770					775					780				
His	Gly	Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Asn	Ile	Gly
785					790					795					800
Gly	Lys	Gln	Ala	Leu	Glu	Ser	Ile	Val	Ala	Gln	Leu	Ser	Arg	Pro	Asp
				805					810					815	
Pro	Ala	Leu	Ala	Ala	Leu	Thr	Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys
			820					825					830		
Leu	Gly	Pro	Lys	Lys	Lys	Arg	Lys	Val							
		835					840								

<210> 27

<211> 2523

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 27

atgctgattg gctatgtaag ggtatcaaca aatgaccaga atacagacct gcaacgaaac	60
gctcttgttt gtgcaggatg tgaacaaata ttggaagata aattaagcgg aacaaggaca	120
gaccgaccgg gattaaaacg cgctttaaag cgccttcaaa aaggtagacac actggttgtc	180
tggaaactgg atcgccctcg gcgaagcatg aaacatttga tttctctcgt aggggaatta	240
cgagagcgag ggattaattt tcgcagtctt actgacagta ttgatacgtc atctccaatg	300
gggcgttttt tcttctacgt tatgggtgcc ctggctgaaa tggaaacgaga actaattatc	360
gagcgaacga tggctggact tgctgccgcc agaaataaag gccgtattgg aggtcgcccg	420
cgtaaactcg ggtctggatc ccccgccggg ccgccgcgcg ccaagccggc cccgcgacgg	480
cgtgctgcgc aacctccga cgcttcgccg gccgcgcagg tggatctacg cacgctcggc	540
tacagtccgc agcagcaaga gaagatcaaa ccgaagggtg gttcgacagt ggccgagcac	600
cacgaggcac tgggtggcca tgggtttaca cacgcgcaca tcgttgcgct cagccaacac	660
ccggcagcgt tagggaccgt cgctgtcaag tatcagcaca taatcacggc gttgccagag	720
gcgacacacg aagacatcgt tggcgtcggc aaacagtggc ccggcgcacg cgccctggag	780
gccttgctca cggatgcggg ggagttgaga ggtccgccgt tacagttgga cacaggccaa	840
cttgtgaaga ttgcaaaacg tggcggcgtg accgcaatgg aggcagtga tgcctcgcgc	900
aatgcactga cgggtgcccc cctggagctg actccggacc aagtgggtggc tatcgccagc	960
aacattggcg gcaagcaagc gctcgaaacg gtgcagcggc tgttgccggc gctgtgccag	1020
gaccatggcc tgactccgga ccaagtggcg gctatcgcca gcaacgggtg cggcaagcaa	1080
gcgctcgaaa cgggtgcagc gctgttgccg gtgctgtgcc aggacctgg cctgactccg	1140
gaccaagtgg tggctatcgc cagcaacatt ggcggcaagc aagcgctcga aacgggtgcag	1200
cggctgttgc cgggtgctgt ccaggacct ggcctgactc cggaccaagt ggtggctatc	1260
gccagcaaca ttggcgga gcaagcgctc gaaacgggtg agcggctgtt gccgggtgctg	1320
tgccaggacc atggcctgac tccggaccaa gtgggtggcta tcgccagcaa cattggcggc	1380
aagcaagcgc tcgaaacggc gcagcggctg ttgccgggtg tgtgccagga ccatggcctg	1440
actccggacc aagtgggtggc tatcgccagc cacgatggcg gcaagcaagc gctcgaaacg	1500
gtgcagcggc tgttgccggc gctgtgccag gaccatggcc tgactccgga ccaagtggcg	1560
gctatcgcca gccacgatgg cggcaagcaa gcgctcgaaa cgggtgcagc gctgttgccg	1620
gtgctgtgcc aggacctgg cctgactccg gaccaagtgg tggctatcgc cagccacgat	1680
ggcggcaagc aagcgctcga aacgggtgcag cggctgttgc cgggtgctgt ccaggacct	1740
ggcctgactc cggaccaagt ggtggctatc gccagccacg atggcgga gcaagcgctc	1800
gaaacgggtg agcggctgtt gccgggtgctg tgccaggacc atggcctgac tccggaccaa	1860

```

gtggtggcta tcgccagcca cgatggcggc aagcaagcgc tcgaaacggt gcagcggctg 1920
ttgccggtgc tgtgccagga ccatggcctg accccggacc aagtgggtggc tatcgccagc 1980
aacggtggcg gcaagcaagc gctcgaaacg gtgcagcggc tgttgccggt gctgtgccag 2040
gaccatggcc tgactccgga ccaagtgggtg gctatcgcca gccacgatgg cggcaagcaa 2100
gcgctcgaaa cgggtgcagcg gctgttgccg gtgctgtgcc aggaccatgg cctgactccg 2160
gaccaagtgg tggctatcgc cagccacgat ggcggcaagc aagcgctcga aacggtgcag 2220
cggctgttgc cgggtgctgtg ccaggaccat ggcccgaccc cggaccaagt ggtggctatc 2280
gccagcaaca ttggcggcaa gcaagcgctc gaaacgggtgc agcggctgtt gccgggtgctg 2340
tgccaggacc atggcctgac cccggaccaa gtggtggcta tcgccagcaa cattggcggc 2400
aagcaagcgc tcgaaagcat tgtggcccag ctgagccggc ctgatccggc gttggccgcg 2460
ttgaccaacg accacctcgt cgccttggcc tgcctcggcc ccaagaagaa gcgcaaggtg 2520
tag 2523

```

<210> 28

<211> 841

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 28

```

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1           5           10           15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
          20           25           30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
          50           55           60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65           70           75           80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
          85           90           95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
          100          105          110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
          115          120          125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Arg Lys Ser Gly
          130          135          140

Ser Gly Ser Pro Ala Leu Arg Pro Pro Arg Ala Lys Pro Ala Pro Arg
145          150          155          160

```

```

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

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

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

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

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

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

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

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

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

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

Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
      325                      330                      335

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

Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
      355                      360                      365

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
      370                      375                      380

Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val
      385                      390                      395                      400

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

Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
      420                      425                      430

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
      435                      440                      445

Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
      450                      455                      460

```

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 465 470 475 480

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
 485 490 495

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 500 505 510

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly
 515 520 525

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

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His
 545 550 555 560

Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 565 570 575

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 580 585 590

Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 595 600 605

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 610 615 620

Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 625 630 635 640

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 645 650 655

Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val
 660 665 670

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

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu
 690 695 700

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 705 710 715 720

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

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 740 745 750

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys
 755 760 765

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 770 775 780

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly
785 790 795 800

Gly Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp
805 810 815

Pro Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys
820 825 830

Leu Gly Pro Lys Lys Lys Arg Lys Val
835 840

<210> 29

<211> 134

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 29

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
1 5 10 15

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
20 25 30

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
35 40 45

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
50 55 60

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp
65 70 75 80

Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
85 90 95

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
100 105 110

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
115 120 125

Ala Leu Thr Gly Ala Pro
130

<210> 30

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 30

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

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

Leu Asn

<210> 31

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 31

Thr	His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp	Ser	Gly	Ala	Arg
1				5					10					15	

Ala	Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	Arg	Gly
			20					25					30

<210> 32

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (12)..(13)

<223> Xaa can be any naturally occurring amino acid

<400> 32

Leu	Thr	Pro	Asp	Gln	Val	Val	Ala	Ile	Ala	Ser	Xaa	Xaa	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> 33

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 33

Asp	Ile	Val	Gly	Val	Lys	Gln	Trp	Ser	Gly	Ala	Arg	Ala	Leu	Glu
1				5					10					15

<210> 34

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 34

Asp	Ile	Val	Gly	Val	Lys	Ser	Arg	Ser	Gly	Ala	Arg	Ala	Leu	Glu
1				5					10					15

<210> 35

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 35

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

<210> 36

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 36

Asp	Ile	Val	Gly	Val	Lys	Tyr	His	Gly	Ala	Arg	Ala	Leu	Glu
1				5					10				

<210> 37

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 37

ataaacccccc tccaaccagg c 21

<210> 38
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 38
aagaaggctct tcattacacc tgcagctctc attttccata cagtcagtat caattctgga 60
agaatttcca g 71

<210> 39
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<220>
<221> misc_feature
<222> (1)..(3)
<223> Xaa can be any naturally occurring amino acid

<400> 39
Xaa Xaa Xaa Ser Gly Ala Arg
1 5

<210> 40
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 40
gtcttcatta cacctgca 18

<210> 41
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 41

cttcattaca cctgcagc 18

<210> 42

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 42

ttcattacac ctgcagct 18

<210> 43

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 43

acctgcagct ctcatttt 18

<210> 44

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 44

gtcagtcata gttaag 16

<210> 45

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 45

tcagtcatag ttaagacc 18

<210> 46

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 46

tcatagttaa gaccttctt 19

<210> 47

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 47

agttaagacc ttcttaa 17

<210> 48

<211> 159

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (104)..(107)

<223> Xaa can be any naturally occurring amino acid

<400> 48

Pro	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala	Pro	Arg	Arg	Arg	Ala	Ala	Gln
1				5					10					15	

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

Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser	Thr
		35					40					45			

Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His	Ala
	50					55					60				

His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro	Ala	Ala	Leu	Gly	Thr	Val	Ala
65					70					75					80

Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	Leu	Pro	Glu	Ala	Thr	His	Glu
				85					90					95	

Asp	Ile	Val	Gly	Val	Gly	Lys	Xaa	Xaa	Xaa	Xaa	Ala	Arg	Ala	Leu	Glu
			100				105						110		

Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln Leu
115 120 125

Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr Ala
130 135 140

Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
145 150 155

<210> 49

<211> 134

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (79)..(82)

<223> Xaa can be any naturally occurring amino acid

<400> 49

Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile
1 5 10 15

Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val
20 25 30

Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His Pro
35 40 45

Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala
50 55 60

Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Xaa Xaa
65 70 75 80

Xaa Xaa Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu
85 90 95

Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala
100 105 110

Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn
115 120 125

Ala Leu Thr Gly Ala Pro
130

<210> 50

<211> 134

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 50

Val	Asp	Leu	Arg	Thr	Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile
1				5					10					15	

Lys	Pro	Lys	Val	Arg	Ser	Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val
			20					25					30		

Gly	His	Gly	Phe	Thr	His	Ala	His	Ile	Val	Ala	Leu	Ser	Gln	His	Pro
		35					40					45			

Ala	Ala	Leu	Gly	Thr	Val	Ala	Val	Lys	Tyr	Gln	Asp	Met	Ile	Ala	Ala
	50					55					60				

Leu	Pro	Glu	Ala	Thr	His	Glu	Ala	Ile	Val	Gly	Val	Gly	Lys	Gln	Trp
65					70					75					80

Ser	Gly	Ala	Arg	Ala	Leu	Glu	Ala	Leu	Leu	Thr	Val	Ala	Gly	Glu	Leu
				85					90					95	

Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Leu	Lys	Ile	Ala
			100					105					110		

Lys	Arg	Gly	Gly	Val	Thr	Ala	Val	Glu	Ala	Val	His	Ala	Trp	Arg	Asn
		115					120					125			

Ala	Leu	Thr	Gly	Ala	Pro
	130				

<210> 51

<211> 80

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 51

Asn	Asp	His	Leu	Val	Ala	Leu	Ala	Cys	Leu	Gly	Gly	Arg	Pro	Ala	Met
1				5					10					15	

Asp	Ala	Val	Lys	Lys	Gly	Leu	Pro	His	Ala	Pro	Glu	Leu	Ile	Arg	Arg
			20					25					30		

Val	Asn	Arg	Arg	Ile	Gly	Glu	Arg	Thr	Ser	His	Arg	Val	Ala	Asp	Tyr
		35					40					45			

Ala	Gln	Val	Val	Arg	Val	Leu	Glu	Phe	Phe	Gln	Cys	His	Ser	His	Pro
	50					55					60				

Ala	Tyr	Ala	Phe	Asp	Glu	Ala	Met	Thr	Gln	Phe	Gly	Met	Ser	Gly	Gln
65					70					75					80

<210> 52

<211> 46

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 52

Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met
1 5 10 15

Asp Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg
20 25 30

Val Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala
35 40 45

<210> 53

$\langle 211 \rangle$ 10

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 53

Asn Asp His Leu Val Ala Leu Ala Cys Leu
1 5 10

<210> 54

<211> 135

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 54

Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys
1 5 10 15

Ile Lys Pro Lys Val Arg Ser Thr Val Ala Gln His His Glu Ala Leu
20 25 30

Val Gly His Gly Phe Thr His Ala His Ile Val Ala Leu Ser Gln His
35 40 45

Pro Ala Ala Leu Gly Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr
50 55 60

Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Ser
65 70 75 80

Arg Ser Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu

				85						90						95			
Leu	Arg	Gly	Pro	Pro	Leu	Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile				
			100					105					110						
Ala	Lys	Arg	Gly	Gly	Val	Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg				
			115				120					125							
Asn	Ala	Leu	Thr	Gly	Ala	Pro													
		130				135													

<210> 55

<211> 132

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 55

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

<210> 56

<211> 180

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 56

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala

100

105

110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Leu Thr
 130 135 140

Lys Ala Glu Trp Glu Gln Ala Gly Arg Leu Leu Ala Gln Gly Ile Pro
 145 150 155 160

Arg Lys Gln Val Ala Leu Ile Tyr Asp Val Ala Leu Ser Thr Leu Tyr
 165 170 175

Lys Lys His Pro
 180

<210> 57

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 57

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60
 Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80
 Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95
 Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110
 Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
 115 120 125
 Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
 130 135 140

<210> 58

<211> 143

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 58

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

<210> 59

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 59

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp

1 5 10 15

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

<210> 60

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 60

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

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Val Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
130 135 140

<210> 61

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 61

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Leu Ser Ile Leu Glu Arg Pro Met Ala Gly His Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
130 135 140

<210> 62

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 62

```

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1      5      10      15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
      20      25      30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
      50      55      60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65      70      75      80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
      85      90      95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
      100      105      110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Ser Ala Gly Arg Ala
      115      120      125

Ala Ala Ile Asn Lys Gly Arg Ile Met Gly Arg Pro Arg Lys Ser Gly
      130      135      140

```

<210> 63

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 63

```

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1      5      10      15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
      20      25      30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
      50      55      60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65      70      75      80

```

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Thr Ser Gly Asn Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp
180 185 190

Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Thr Ser Gly Glu Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 64

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 64

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

His Arg Thr Thr Leu Thr Asn His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp
180 185 190

Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 65

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 65

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

-- -- --

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Gln Ser Gly Asp Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His
180 185 190

Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Arg Ser Asp Glu Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 66

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 66

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu

```

      - 20      - 25      - 30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
 35      40      45

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50      55      60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65      70      75      80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
      85      90      95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
      100      105      110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
      115      120      125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
      130      135      140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145      150      155      160

Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
      165      170      175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Lys Asp Asn
      180      185      190

Leu Lys Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
      195      200      205

Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Glu Leu Val Arg His
      210      215      220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225      230      235      240

Lys Ser Phe Ser Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His
      245      250      255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
      260      265

```

<210> 67

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 67

```

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1      5      10      15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20      25      30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50      55      60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65      70      75      80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85      90      95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100      105      110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115      120      125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130      135      140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145      150      155      160

Thr Thr Gly Asn Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys
165      170      175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly Ala
180      185      190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195      200      205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His
210      215      220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225      230      235      240

Lys Ser Phe Ser Arg Ser Asp His Leu Thr Asn His Gln Arg Thr His
245      250      255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260      265

```

<210> 68

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 68

Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp
1				5					10					15	

Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu
			20					25					30		

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

Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp
	50					55					60				

Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu
65					70					75					80

Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr
				85					90					95	

Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala
			100					105					110		

Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala
		115					120					125			

Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Pro	Lys	Ser	Gly
	130					135					140				

Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser
145					150					155					160

Arg	Lys	Asp	Asn	Leu	Lys	Asn	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys
			165					170						175	

Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Arg	Ser	Asp	His
			180					185						190	

Leu	Thr	Asn	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys
		195					200					205			

Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Asp	Pro	Gly	Asn	Leu	Val	Arg	His
	210					215					220				

Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

225							230								240
-----	--	--	--	--	--	--	-----	--	--	--	--	--	--	--	-----

Lys	Ser	Phe	Ser	Arg	Lys	Asp	Asn	Leu	Lys	Asn	His	Gln	Arg	Thr	His
				245					250					255	

Thr	Gly	Lys	Lys	Thr	Ser	Gly	Gln	Ala	Gly	Gln					
			260					265							

<210> 69

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 69

Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp
1			5						10					15	

Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu
			20					25					30		

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

Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp
	50					55					60				

Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu
65					70					75					80

Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr
				85					90					95	

Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala
			100					105					110		

Glu	Met	Glu	Arg	Glu	Leu	Ile	Leu	Glu	Arg	Val	Met	Ala	Gly	Ile	Ala
		115					120					125			

Ala	Ala	Arg	Asn	Lys	Gly	Arg	Arg	Trp	Gly	Arg	Pro	Pro	Lys	Ser	Gly
	130					135					140				

Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser
145					150					155					160

Gln	Arg	Ala	Asn	Leu	Arg	Ala	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys
				165					170					175	

Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Ser	Ser
			180					185					190		

Leu	Val	Arg	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys
		195					200					205			

Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Thr	Thr	Gly	Asn	Leu	Thr	Val	His
	210					215					220				

Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly
225					230					235					240

Lys	Ser	Phe	Ser	Gln	Arg	Ala	His	Leu	Glu	Arg	His	Gln	Arg	Thr	His
				245					250					255	

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 70

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 70

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Arg Asp Glu
180 185 190

Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly

225					230					235					240
Lys	Ser	Phe	Ser	Gln	Arg	Ala	His	Leu	Glu	Arg	His	Gln	Arg	Thr	His
				245					250					255	

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 71

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 71

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Arg Arg Asp Glu Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His
180 185 190

Leu Thr Asn His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 72

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 72

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Thr Ser Gly Ser Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Lys
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Gly Asp Leu Arg Arg His
 210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240

Lys Ser Phe Ser Thr Ser Gly Glu Leu Val Arg His Gln Arg Thr His
 245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265

<210> 73

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 73

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
 100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Ser Ala Gly Arg Ala
 115 120 125

Ala Ala Ile Asn Lys Gly Arg Ile Met Gly Arg Pro Arg Lys Ser Gly
 130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160

Gln Leu Ala His Leu Arg Ala His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His

180

185

190

Leu Arg Ala His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly His Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Asp Ser Gly Asn Leu Arg Val His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 74

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 74

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Thr Gly Asn
180 185 190

Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Asp Ser Gly Asn Leu Arg Val His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 75

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 75

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Thr His Leu Asp Leu Ile Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Thr Gly Asn
180 185 190

Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Arg Ser Asp Asn Leu Val Arg His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 76

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 76

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Ile Ala
 115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Ser Gly
 130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 145 150 155 160

Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
 165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Arg Asp Glu
 180 185 190

Leu Asn Val His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His
 210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
 225 230 235 240

Lys Ser Phe Ser Arg Ser Asp His Leu Thr Asn His Gln Arg Thr His
 245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
 260 265

<210> 77

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 77

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
 1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
 20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
 50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
 65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
 85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Gln Arg Ala His Leu Glu Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Glu Leu Val Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser His Lys Asn Ala Leu Gln Asn His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 78

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 78

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

```

      ~~~~~
Arg  Glu  Arg  Gly  Ile  Asn  Phe  Arg  Ser  Leu  Thr  Asp  Ser  Ile  Asp  Thr
      85                      90                      95

Ser  Ser  Pro  Met  Gly  Arg  Phe  Phe  Phe  Tyr  Val  Met  Gly  Ala  Leu  Ala
      100                      105                      110

Glu  Met  Glu  Arg  Glu  Leu  Ile  Ile  Glu  Arg  Thr  Met  Ala  Gly  Leu  Ala
      115                      120                      125

Ala  Ala  Arg  Asn  Lys  Gly  Arg  Ile  Gly  Gly  Arg  Pro  Pro  Lys  Ser  Gly
      130                      135                      140

Thr  Gly  Glu  Lys  Pro  Tyr  Lys  Cys  Pro  Glu  Cys  Gly  Lys  Ser  Phe  Ser
145                      150                      155                      160

Arg  Arg  Asp  Glu  Leu  Asn  Val  His  Gln  Arg  Thr  His  Thr  Gly  Glu  Lys
      165                      170                      175

Pro  Tyr  Lys  Cys  Pro  Glu  Cys  Gly  Lys  Ser  Phe  Ser  Gln  Ser  Ser  Asn
      180                      185                      190

Leu  Val  Arg  His  Gln  Arg  Thr  His  Thr  Gly  Glu  Lys  Pro  Tyr  Lys  Cys
      195                      200                      205

Pro  Glu  Cys  Gly  Lys  Ser  Phe  Ser  Gln  Ser  Ser  Ser  Leu  Val  Arg  His
      210                      215                      220

Gln  Arg  Thr  His  Thr  Gly  Glu  Lys  Pro  Tyr  Lys  Cys  Pro  Glu  Cys  Gly
225                      230                      235                      240

Lys  Ser  Phe  Ser  Thr  Thr  Gly  Asn  Leu  Thr  Val  His  Gln  Arg  Thr  His
      245                      250                      255

Thr  Gly  Lys  Lys  Thr  Ser  Gly  Gln  Ala  Gly  Gln
      260                      265

```

<210> 79

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 79

```

Met  Leu  Ile  Gly  Tyr  Val  Arg  Val  Ser  Thr  Asn  Asp  Gln  Asn  Thr  Asp
1                      5                      10                      15

Leu  Gln  Arg  Asn  Ala  Leu  Val  Cys  Ala  Gly  Cys  Glu  Gln  Ile  Phe  Glu
      20                      25                      30

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

```

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Leu Glu Arg Val Met Ala Gly Ile Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Arg Trp Gly Arg Pro Pro Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Thr Thr Gly Asn Leu Thr Val His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu Glu Arg His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Lys Ser Ser Leu Ile Ala His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 80

<211> 267

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 80

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Ser Ala Gly Arg Ala
115 120 125

Ala Ala Ile Asn Lys Gly Arg Ile Met Gly Arg Pro Arg Lys Ser Gly
130 135 140

Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
145 150 155 160

Asp Pro Gly Ala Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
165 170 175

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser
180 185 190

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
195 200 205

Pro Glu Cys Gly Lys Ser Phe Ser Gln Leu Ala His Leu Arg Ala His
210 215 220

Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly
225 230 235 240

Lys Ser Phe Ser Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His
245 250 255

Thr Gly Lys Lys Thr Ser Gly Gln Ala Gly Gln
260 265

<210> 81

<211> 938

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 81

Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr
1 5 10 15

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
 20 25 30

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
 35 40 45

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
 50 55 60

Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr
 65 70 75 80

His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala
 85 90 95

Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu
 100 105 110

Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val
 115 120 125

Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala
 130 135 140

Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly
 145 150 155 160

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
 165 170 175

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

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

Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile
 210 215 220

Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 225 230 235 240

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val
 245 250 255

Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
 260 265 270

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
 275 280 285

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

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
 305 310 315 320

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

640	650	660	670
Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His	645	650	655
Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly	660	665	670
Lys Gln Ala Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro	675	680	685
Ala Leu Ala Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu	690	695	700
Gly Gly Arg Pro Ala Met Asp Ala Val Lys Lys Gly Leu Pro His Ala			
705	710	715	720
Pro Glu Leu Ile Arg Arg Val Asn Arg Arg Ile Gly Glu Arg Thr Ser	725	730	735
His Arg Val Ala Gly Ser Gln Leu Val Lys Ser Glu Leu Glu Glu Lys	740	745	750
Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile	755	760	765
Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu	770	775	780
Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys	785	790	800
His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly	805	810	815
Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly	820	825	830
Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val	835	840	845
Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp	850	855	860
Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser	865	870	875
Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His	885	890	895
Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile	900	905	910
Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg	915	920	925

Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
930 935

<210> 82

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 82

Asn Gly Asn Asn His Asp Asn Gly His Asp Asn Ile Asn Gly Asn Gly
1 5 10 15

Asn Ile His Asp Asn Ile His Asp His Asp Asn Gly Asn Asn His Asp
20 25 30

Asn Ile

<210> 83

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 83

His Asp Asn Ile Asn Asn Asn Gly His Asp Asn Ile Asn Asn Asn Gly
1 5 10 15

Asn Ile Asn Gly His Asp Asn Ile Asn Ile Asn Gly Asn Gly
20 25 30

<210> 84

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 84

His Asp His Asp Asn Gly Asn Asn His Asp Asn Ile Asn Asn His Asp
1 5 10 15

Asn Gly His Asp Asn Gly His Asp Asn Ile Asn Gly Asn Gly Asn Gly
20 25 30

Asn Gly

<210> 85

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 85

Asn	Ile	Asn	Gly	Asn	Gly	His	Asp	Asn	Gly	Asn	Gly	His	Asp	His	Asp
1				5					10					15	

Asn	Ile	Asn	Asn	Asn	Ile	Asn	Gly	Asn	Gly	Asn	Asn	Asn	Ile
			20					25					30

<210> 86

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 86

His	Asp	Asn	Ile	Asn	Asn	Asn	Ile	Asn	Ile	Asn	Gly	Asn	Gly	Asn	Asn
1				5					10					15	

Asn	Ile	Asn	Gly	Asn	Ile	His	Asp	Asn	Gly	Asn	Asn	Asn	Ile	His	Asp
			20					25					30		

Asn Gly

<210> 87

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 87

Asn	Gly	His	Asp	Asn	Ile	Asn	Gly	Asn	Gly	Asn	Ile	His	Asp	Asn	Ile
1				5					10					15	

His	Asp	His	Asp	Asn	Gly	Asn	Asn	His	Asp	Asn	Ile	Asn	Asn	His	Asp
			20					25					30		

<210> 88

<211> 36

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 88

His Asp Asn Gly Asn Gly His Asp His Asp Asn Ile Asn Asn Asn Ile
 1 5 10 15

Asn Ile Asn Gly Asn Gly Asn Asn Asn Ile Asn Gly Asn Ile His Asp
 20 25 30

Asn Gly Asn Asn
 35

<210> 89

<211> 432

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 89

atgagatctc ctaagaaaaa gaggaagatg gtggacttga ggacactcgg ttattcgcaa 60
 cagcaacagg agaaaatcaa gcctaaggtc aggagcaccg tcgcgcaaca ccacgaggcg 120
 cttgtggggc atggcttcac tcatgcgcat attgtcgcgc ttccacagca ccctgcggcg 180
 cttgggacgg tggctgtcaa ataccaagat atgattgcgg ccctgcccga agccacgcac 240
 gaggcaattg taggggtcgg taaacagtgg tcgggagcgc gagcacttga ggcgctgctg 300
 actgtggcgg gtgagcttag ggggcctccg ctccagctcg acaccgggca gctgctgaag 360
 atcgcgaaga gagggggagt aacagcggta gaggcagtgc atgcatcgcg caatgcactg 420
 acgggtgccc cc 432

<210> 90

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 90

Met Arg Ser Pro Lys Lys Lys Arg Lys Met Val Asp Leu Arg Thr Leu
 1 5 10 15

Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg Ser
 20 25 30

Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr His
 35 40 45

Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr Val
 50 55 60

Ala Val Lys Tyr Gln Asp Met Ile Ala Ala Leu Pro Glu Ala Thr His
 65 70 75 80

Glu Ala Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg Ala Leu
85 90 95

Glu Ala Leu Leu Thr Val Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln
100 105 110

Leu Asp Thr Gly Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly Val Thr
115 120 125

Ala Val Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
130 135 140

<210> 91

<211> 432

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 91

```
atgagatctc ctaagaaaa gaggaaggtg caggtggatc tacgcacgct cggctacagt      60
cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacgag      120
gcactggtgg gccatgggtt tacacacgcg cacatcgttg cgctcagcca acacccggca      180
gcgttaggga ccgtcgtctg cacgtatcag cacataatca cggcgttgcc agaggcgaca      240
cacgaagaca tcgttggcgt cggcaaatat catggggcac gcgctctgga ggccttgctc      300
acggatgcgg gggagttgag aggtccgcgc ttacagttgg acacaggcca acttgtgaag      360
attgcaaaac gtggcggcgt gaccgcaatg gaggcagtgc atgcatcgcg caatgcactg      420
acgggtgccc cc                                     432
```

<210> 92

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 92

```
Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr
1          5          10          15

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
20          25          30

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
35          40          45

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
50          55          60
```

Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr
65 70 75 80

His Glu Asp Ile Val Gly Val Gly Lys Tyr His Gly Ala Arg Ala Leu
85 90 95

Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu Gln
100 105 110

Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val Thr
115 120 125

Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro
130 135 140

<210> 93

<211> 435

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 93

atgagatctc ctaagaaaa gaggaaggtg caggtggatc tacgcacgct cggctacagt	60
cagcagcagc aagagaagat caaaccgaag gtgcgttcga cagtggcgca gcaccacgag	120
gcactggtgg gccatgggtt tacacacgcg cacatcggtg cgctcagcca acaccggca	180
gcgttaggga ccgtcgctgt cacgtatcag cacataatca cggcgttgcc agaggcgaca	240
cacgaagaca tcgttggcgt cggcaaatcg cggtcggggg cacgcgctct ggaggccttg	300
ctcacggatg cgggggagtt gagaggtccg ccgttacagt tggacacagg ccaacttgtg	360
aagattgcaa aacgtggcgg cgtgaccgca atggaggcag tgcattgcac gcgcaatgca	420
ctgacgggtg ccccc	435

<210> 94

<211> 145

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 94

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

50		55		60											
Val	Ala	Val	Thr	Tyr	Gln	His	Ile	Ile	Thr	Ala	Leu	Pro	Glu	Ala	Thr
65					70					75					80
His	Glu	Asp	Ile	Val	Gly	Val	Gly	Lys	Ser	Arg	Ser	Gly	Ala	Arg	Ala
			85						90					95	
Leu	Glu	Ala	Leu	Leu	Thr	Asp	Ala	Gly	Glu	Leu	Arg	Gly	Pro	Pro	Leu
			100					105					110		
Gln	Leu	Asp	Thr	Gly	Gln	Leu	Val	Lys	Ile	Ala	Lys	Arg	Gly	Gly	Val
		115					120					125			
Thr	Ala	Met	Glu	Ala	Val	His	Ala	Ser	Arg	Asn	Ala	Leu	Thr	Gly	Ala
	130					135					140				

Pro
145

<210> 95

<211> 435

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 95

atgagatctc ctaagaaaaa gaggaaggtg caggtggatc tacgcacgct cggctacagt	60
cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacggg	120
gcactgggtg gccatgggtt tacacacgcg cacatcggtg cgctcagcca acaccgggca	180
gcgttaggga ccgtcgctgt cacgtatcag cacataatca cggcgttgcc agaggcgaca	240
cacgaagaca tcgttggcgt cggcaaaccg ggggctgggt cacgcgctct ggaggccttg	300
ctcacggatg cgggggagtt gagaggtccg ccgttacagt tggacacagg ccaacttgtg	360
aagattgcaa aacgtggcgg cgtgaccgca atggaggcag tgcattgcat gcgcaatgca	420
ctgacgggtg ccccc	435

<210> 96

<211> 145

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 96

Met	Arg	Ser	Pro	Lys	Lys	Lys	Arg	Lys	Val	Gln	Val	Asp	Leu	Arg	Thr
1				5					10					15	
Leu	Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg

20

25

30

Ser Thr Val Ala Gln His His Gly Ala Leu Val Gly His Gly Phe Thr
354045

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
505560

Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr
65707580

His Glu Asp Ile Val Gly Val Gly Lys Arg Gly Ala Gly Ala Arg Ala
859095

Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu
100105110

Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val
115120125

Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala
130135140

Pro
145

- <210> 97
- <211> 435
- <212> DNA
- <213> Artificial Sequence
- <220>
- <223> Synthetic construct

<400> 97

atgagatctc ctaagaaaa gaggaagggtg caggtggatc tacgcacgct cggctacagt60

cagcagcagc aagagaagat caaacccaag gtgcgttcga cagtggcgca gcaccacgag120

gcactgggtg gccatgggtt tacacacgcg cacatcggtg cgctcagcca acacccggca180

gcgttaggga ccgtcgctgt cacgtatcag cacataatca cggcggttgcc agaggcgaca240

cacgaagaca tcgttggcgt cggcaaacag tggtcggcg caccgcgcct ggaggccttg300

ctcacggatg cgggggagtt gagaggccg ccgttacagt tggacacagg ccaacttgtg360

aagattgcaa aacgtggcgg cgtgaccgca atggaggcag tgcatgcatc gcgcaatgca420

ctgacgggtg ccccc435

- <210> 98
- <211> 1080
- <212> PRT
- <213> Artificial Sequence
- <220>
- <223> Synthetic construct

<220>

<221> MISC_FEATURE

<222> (498)..(498)

<223> Xaa is Q, S, R or Y

<220>

<221> MISC_FEATURE

<222> (499)..(499)

<223> Xaa is W, R or G

<220>

<221> MISC_FEATURE

<222> (500)..(500)

<223> Xaa is S, A or H

<400> 98

Met	Lys	Ile	Glu	Glu	Gly	Lys	Leu	Val	Ile	Trp	Ile	Asn	Gly	Asp	Lys
1			5					10					15		

Gly	Tyr	Asn	Gly	Leu	Ala	Glu	Val	Gly	Lys	Lys	Phe	Glu	Lys	Asp	Thr
		20					25					30			

Gly	Ile	Lys	Val	Thr	Val	Glu	His	Pro	Asp	Lys	Leu	Glu	Glu	Lys	Phe
	35					40					45				

Pro	Gln	Val	Ala	Ala	Thr	Gly	Asp	Gly	Pro	Asp	Ile	Ile	Phe	Trp	Ala
	50					55					60				

His	Asp	Arg	Phe	Gly	Gly	Tyr	Ala	Gln	Ser	Gly	Leu	Leu	Ala	Glu	Ile
65				70					75					80	

Thr	Pro	Asp	Lys	Ala	Phe	Gln	Asp	Lys	Leu	Tyr	Pro	Phe	Thr	Trp	Asp
			85						90					95	

Ala	Val	Arg	Tyr	Asn	Gly	Lys	Leu	Ile	Ala	Tyr	Pro	Ile	Ala	Val	Glu
		100						105					110		

Ala	Leu	Ser	Leu	Ile	Tyr	Asn	Lys	Asp	Leu	Leu	Pro	Asn	Pro	Pro	Lys
	115						120					125			

Thr	Trp	Glu	Glu	Ile	Pro	Ala	Leu	Asp	Lys	Glu	Leu	Lys	Ala	Lys	Gly
	130					135					140				

Lys	Ser	Ala	Leu	Met	Phe	Asn	Leu	Gln	Glu	Pro	Tyr	Phe	Thr	Trp	Pro
145				150					155						160

Leu	Ile	Ala	Ala	Asp	Gly	Gly	Tyr	Ala	Phe	Lys	Tyr	Glu	Asn	Gly	Lys
				165				170						175	

Tyr	Asp	Ile	Lys	Asp	Val	Gly	Val	Asp	Asn	Ala	Gly	Ala	Lys	Ala	Gly
			180					185					190		

Leu	Thr	Phe	Leu	Val	Asp	Leu	Ile	Lys	Asn	Lys	His	Met	Asn	Ala	Asp
	195						200					205			

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

Gly Glu Leu Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu
 515 520 525

Val Lys Ile Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His
 530 535 540

Ala Ser Arg Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp
 545 550 555 560

Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
 565 570 575

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 580 585 590

Pro Asp Gln Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala
 595 600 605

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 610 615 620

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys
 625 630 635 640

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 645 650 655

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly
 660 665 670

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

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn
 690 695 700

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 705 710 715 720

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 725 730 735

Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
 740 745 750

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala
 755 760 765

Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
 770 775 780

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 785 790 795 800

Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val
 805 810 815

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
 820 825 830
 Gln Val Val Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala Leu Glu
 835 840 845
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
 850 855 860
 Pro Asp Gln Val Val Ala Ile Val Ser His Asp Gly Gly Lys Gln Ala
 865 870 875 880
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
 885 890 895
 Leu Thr Pro Asp Gln Val Val Ala Ile Val Ser Asn Gly Gly Gly Lys
 900 905 910
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
 915 920 925
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly
 930 935 940
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
 945 950 955 960
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His
 965 970 975
 Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
 980 985 990
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala
 995 1000 1005
 Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
 1010 1015 1020
 Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
 1025 1030 1035
 Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser
 1040 1045 1050
 Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu
 1055 1060 1065
 Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
 1070 1075 1080

<210> 99

<211> 830

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> MISC_FEATURE

<222> (249)..(249)

<223> Xaa is Q, S, R or Y

<220>

<221> MISC_FEATURE

<222> (250)..(250)

<223> Xaa is W, R or G

<220>

<221> MISC_FEATURE

<222> (251)..(251)

<223> Xaa is S, A or H

<400> 99

Met	Leu	Ile	Gly	Tyr	Val	Arg	Val	Ser	Thr	Asn	Asp	Gln	Asn	Thr	Asp
1				5					10					15	

Leu	Gln	Arg	Asn	Ala	Leu	Val	Cys	Ala	Gly	Cys	Glu	Gln	Ile	Phe	Glu
			20					25					30		

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

Leu	Lys	Arg	Leu	Gln	Lys	Gly	Asp	Thr	Leu	Val	Val	Trp	Lys	Leu	Asp
	50					55					60				

Arg	Leu	Gly	Arg	Ser	Met	Lys	His	Leu	Ile	Ser	Leu	Val	Gly	Glu	Leu
65					70					75					80

Arg	Glu	Arg	Gly	Ile	Asn	Phe	Arg	Ser	Leu	Thr	Asp	Ser	Ile	Asp	Thr
				85					90					95	

Ser	Ser	Pro	Met	Gly	Arg	Phe	Phe	Phe	Tyr	Val	Met	Gly	Ala	Leu	Ala
			100					105					110		

Glu	Met	Glu	Arg	Glu	Leu	Ile	Ile	Glu	Arg	Thr	Met	Ala	Gly	Leu	Ala
		115					120					125			

Ala	Ala	Arg	Asn	Lys	Gly	Arg	Ile	Gly	Gly	Arg	Pro	Pro	Lys	Ser	Gly
	130					135					140				

Ser	Pro	Arg	Pro	Pro	Arg	Ala	Lys	Pro	Ala	Pro	Arg	Arg	Arg	Ala	Ala
145					150					155				160	

Gln	Pro	Ser	Asp	Ala	Ser	Pro	Ala	Ala	Gln	Val	Asp	Leu	Arg	Thr	Leu
			165						170					175	

Gly	Tyr	Ser	Gln	Gln	Gln	Gln	Glu	Lys	Ile	Lys	Pro	Lys	Val	Arg	Ser
			180					185					190		

Thr	Val	Ala	Gln	His	His	Glu	Ala	Leu	Val	Gly	His	Gly	Phe	Thr	His
		195					200					205			

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

```

110 Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala
515                               520                               525

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly
530                               535                               540

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
545                               550                               555                               560

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp
565                               570                               575

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Val Ser His Asp Gly
580                               585                               590

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys
595                               600                               605

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Val Ser His
610                               615                               620

Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val
625                               630                               635                               640

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Val
645                               650                               655

Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu
660                               665                               670

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

Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg
690                               695                               700

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val
705                               710                               715                               720

Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val
725                               730                               735

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp
740                               745                               750

Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu
755                               760                               765

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr
770                               775                               780

Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala
785                               790                               795                               800

Leu Glu Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala
805                               810                               815

```

Ala Leu Thr Asn Asp His Leu Val Ala Leu Ala Cys Leu Gly
820 825 830

<210> 100

<211> 839

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> MISC_FEATURE

<222> (112)..(112)

<223> Xaa is Q, S, R or Y

<220>

<221> MISC_FEATURE

<222> (113)..(113)

<223> Xaa is W, R or R

<220>

<221> MISC_FEATURE

<222> (114)..(114)

<223> Xaa is S, A or H

<400> 100

Met Ala Gln Ala Ala Ser Gly Ser Pro Arg Pro Pro Arg Ala Lys Pro
1 5 10 15

Ala Pro Arg Arg Arg Ala Ala Gln Pro Ser Asp Ala Ser Pro Ala Ala
20 25 30

Gln Val Asp Leu Arg Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys
35 40 45

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

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

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

Ala Leu Pro Glu Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Xaa
100 105 110

Xaa Xaa Gly Ala Arg Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu
115 120 125

Leu Arg Gly Pro Pro Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile
130 135 140

Ala Lys Arg Gly Gly Val Thr Ala Met Glu Ala Val His Ala Ser Arg
145 150 155 160

Asn Ala Leu Thr Gly Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val
165 170 175

Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln
180 185 190

Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln
195 200 205

Val Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr
210 215 220

Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro
225 230 235 240

Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu
245 250 255

Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu
260 265 270

Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys Gln
275 280 285

Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His
290 295 300

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

Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln
325 330 335

Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp
340 345 350

Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu
355 360 365

Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser
370 375 380

His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro
385 390 395 400

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

Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu
420 425 430

Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val
435 440 445

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

Ala Met Thr Gln Phe Gly Met Ser Gly Gln Ala Gly Gln Ala Ser Pro
755 760 765

Lys Lys Lys Arg Lys Val Gly Arg Ala Asp Ala Leu Asp Asp Phe Asp
770 775 780

Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met
785 790 795 800

Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser
805 810 815

Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn Tyr Pro Tyr
820 825 830

Asp Val Pro Asp Tyr Ala Ser
835

<210> 101

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 101

Asn Ile Asn Gly Asn Ile Asn Ile Asn Ile His Asp His Asp His Asp
1 5 10 15

His Asp His Asp Asn Ile His Asp His Asp Asn Ile Asn Ile
20 25 30

<210> 102

<211> 33

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 102

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

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

Asn

<210> 103

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 103

Lys Arg Gly Gly
1

<210> 104

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 104

Leu Asp Tyr Glu
1

<210> 105

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 105

Ile Asn Leu Val
1

<210> 106

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 106

Tyr Ser Lys Lys

1

<210> 107

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 107
Asn Met Ala His
1

<210> 108
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 108
Ser Pro Thr Asn
1

<210> 109
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 109
Ser Asn Thr Arg
1

<210> 110
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 110
Leu Thr Thr Thr
1

<210> 111
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 111
Val Ala Asp Leu
1

<210> 112

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

<220>
<223> Synthetic construct

<400> 112
Met Val Leu Ser
1

<210> 113
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 113
Tyr Asn Gly Arg
1

<210> 114
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 114
Arg Ile Pro Arg
1

<210> 115
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 115
Tyr Ser Lys Ile

1

<210> 116
<211> 4
<212> PRT
<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 116

Leu Thr Gln Tyr
1

<210> 117

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 117

Tyr Leu Ser Lys
1

<210> 118

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 118

Leu Arg Pro Asn
1

<210> 119

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 119

Leu Phe Thr Asn
1

<210> 120

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 120

Leu Leu Thr Asn

Leu Leu Thr Asp
1

<210> 121
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 121
Glu Glu Asp Lys
1

<210> 122
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 122
Val Thr Ala Met
1

<210> 123
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 123
Cys Pro Ser Arg
1

<210> 124
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 124
Leu Thr Arg Val

<210> 125

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

<220>
 <223> Synthetic construct

<400> 125
 Lys Gly Asp Leu
 1

<210> 126
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 126
 Gln Lys Ala Leu
 1

<210> 127
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 127
 Leu Tyr Leu Leu
 1

<210> 128
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 128
 Trp Ile Ser Val
 1

<210> 129
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 129

Gly Asp Gln Val
1

<210> 130

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 130

Cys Pro Ser Arg
1

<210> 131

<211> 145

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (89)..(93)

<223> Xaa can be any naturally occurring amino acid

<400> 131

Met Arg Ser Pro Lys Lys Lys Arg Lys Val Gln Val Asp Leu Arg Thr
1 5 10 15

Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val Arg
20 25 30

Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe Thr
35 40 45

His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly Thr
50 55 60

Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala Thr
65 70 75 80

His Glu Asp Ile Val Gly Val Gly Xaa Xaa Xaa Xaa Xaa Ala Arg Ala
85 90 95

Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro Leu
100 105 110

Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly Val
115 120 125

Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly Ala
130 135 140

Pro
145

<210> 132
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 132
Lys Arg Pro Ala Gly
1 5

<210> 133
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 133
Lys Arg Pro Ser Gly
1 5

<210> 134
<211> 32
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 134
Leu Thr Pro Asp Val Val Ala Ile Ser Asn Asn Gly Gly Lys Gln Ala
1 5 10 15

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

<210> 135
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 135
Ser Asn Asn Gly
1

<210> 136
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 136
Arg Gly Gly Gly
1

<210> 137
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 137
Arg Gly Gly Arg
1

<210> 138
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 138
Arg Gly Val Arg
1

<210> 139
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 139
Lys Gly Gly Gly
1

<210> 140

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 140

Ser Gly Gly Gly
1

<210> 141

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 141

Gly Gly Arg Gly
1

<210> 142

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 142

Ileu Gly Gly Ser
1

<210> 143

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 143

Met Asp Asn Ile
1

<210> 144

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 144
Arg Val Met Ala
1

<210> 145
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 145
Leu Ala Ser Val
1

<210> 146
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 146
Val Gly Thr Gly
1

<210> 147
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 147
Gln Gly Gly Gly
1

<210> 148
<211> 118
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 148	
ttaattaaga gtctagaaat ataaaccccc tccaaccagg tgctaactgt aaaccatggt	60
tttgaattaa cacctgattc aaggaatttt ataatatcta gaaggaattt aaaatgaa	118

<210> 149

<211> 117

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 149

actgacctag agaagcttat ataaaccccc tccaaccagg tgctaatacca aaaccatggt 60

ttacagtttag cacctgggtg gaggggggtt atactgcagt tatttgtaca gttcatc 117

<210> 150

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 150

ttaattaaga gtctagatta gcacctgggt ggaggggggt tataagggtt tggtagcaaa 60

tgtctataaaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 151

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 151

ttaattaaga gtctagatta gcacctgggt ggaggggggt tataagggtt tggtagcaaa 60

tgtctataaaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 152

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 152

actgacctag agaagctttt agcacctggt tggaggggggt ttatagacat ttggtaccaa 60

aaccttataa acccctcca accaggtgct aactgcagtt atttgtacag ttcac 116

<210> 153

<211> 116
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 153
ttaaattaaga gtctagatta gcacctgggt ggaggggggt tatatccaaa accatgggtt 60
acagtataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 154
<211> 116
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 154
actgacctag agaagctttt agcacctgggt tggaggggggt ttatatccaa aacctgggt 60
tacagtataa acccctcca accaggtgct aactgcagtt atttgtacag ttcac 116

<210> 155
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 155
ttaaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttcc aaaacctagg 60
tttacagggt tataaaccct ctccaaccag gtgctaaaga tctaggagga atttaaaatg 120
ag 122

<210> 156
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 156
ttaaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca tccaaaacca 60
tggtttacag ggttcctata aaccctctcc aaccaggtgc taaagatcta ggaggaattt 120
aaaatgag 128

<210> 157
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 157
actgacctag agaagctttt agcacctggg tggagggggg ttatagcaac cctgtaaacc 60
atggtttttg atgaagctat aaacccccct caaccagggt ctaactgcag ttatttgtac 120
agttcatc 128

<210> 158
<211> 140
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 158
ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca gtttcatcca 60
aaaccatggt ttacagggtt ccggttccta taaacccccct ccaaccagggt gctaaagatc 120
taggaggaat ttaaaatgag 140

<210> 159
<211> 140
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 159
ttaattaaga gtctagatta gcacctgggt ggaggggggt tatagcttca gtttcatcca 60
aaaccatggt ttacagggtt ccggttccta taaacccccct ccaaccagggt gctaaagatc 120
taggaggaat ttaaaatgag 140

<210> 160
<211> 110
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 160
ttaattaaga gtctagatta gcacctgggt ggaggggggt tataaaaacc atggtttata 60

taaaccocct ccaaccaggt gctaaagatc taggaggaat ttaaaatgag 110

<210> 161

<211> 110

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 161

actgacctag agaagctttt agcacctggt tggagggggt ttatataaac catgggtttt 60

ataaaccccc tccaaccagg tgctaactgc agttatttgc acagttcatc 110

<210> 162

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 162

ttaattaaga gtctagatta gcacctggtt ggaggggggt tatatccaaa accgggggtt 60

acagtataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 163

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 163

actgacctag agaagctttt agcacctggt tggagggggt ttatactgta aaccccggtt 60

ttggatataa acccctcca accaggtgct aactgcagtt atttgtacag ttcac 116

<210> 164

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 164

ttaattaaga gtctagatta gcacctggtt ggaggggggt tatacgaaat attataaatt 60

atcatataaa cccctccaa ccaggtgcta aagatctagg aggaatttaa aatgag 116

<210> 165

<211> 116

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 165

actgacctag agaagctttt agcacctggt tggaggggggt ttatatgata atttataata 60

tttcgtataa accccctcca accaggtgct aactgcagtt atttgtacag ttcac 116

<210> 166

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 166

ttaattaaga gtctagatta gcacctggtt ggaggggggt tatagcttca tccaaaaccg 60

gggttttacag ggttcctata aacccctcc aaccaggtgc taaagatcta ggaggaattt 120

aaaatgag 128

<210> 167

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 167

actgacctag agaagctttt agcacctggt tggaggggggt ttatagcaac cctgtaaacc 60

ggggtttttg atgaagctat aaacccctc caaccaggtg ctaactgcag ttatttgtac 120

agttcatc 128

<210> 168

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 168

ttaattaaga gtctagatta gcacctggtt ggaggggggt tatagcttca cgaaatatta 60

ttaattatca ggttcctata aacccctcc aaccaggtgc taaagatcta ggaggaattt 120

aaaatgag 128

<210> 169
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 169
actgacctag agaagctttt agcacctggt tggaggggggt ttatagcaac ctgataattt 60
ataatatttc gtgaagctat aaaccccttc caaccaggtg ctaactgcag ttatttgtac 120
agttcatc 128

<210> 170
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 170
ttaattaaga gagatcttta gcacctggtt ggaggggggt tatagcttca tccaaaacca 60
tggtttacag ggttcctata aaccccttc aaccaggtgc taagcgatct gcatctcaat 120
tagtcagc 128

<210> 171
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 171
actgacctag agaagctttt agcacctggt tggaggggggt ttatagcaac cctgtaaacc 60
atggtttttg atgaagctat aaaccccttc caaccaggtg ctaatttgca aaagcctagg 120
cctccaaa 128

<210> 172
<211> 107
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 172

ttaattaaga gagatctgcg ggaggcgtgt ccaaaacat ggtttacagg gttcctataa	60
acccctcca accagggtgct aagcgatctg catctcaatt agtcagc	107

<210> 173

<211> 107

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 173

actgacctag agaagctttt agcacctggt tggagggggt ttatagcaac cctgtaaacc	60
atggtttttg acacgcctcc cgctttgcaa aagcctaggc ctccaaa	107

<210> 174

<211> 140

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 174

ttaattaaga gagatcttta gcacctggtt ggagggggtt tatagcttca gttcatcca	60
aaacctggt ttacagggtt ccggttcta taaacccct ccaaccagggt gctaagcgat	120
ctgcatctca attagtcagc	140

<210> 175

<211> 140

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 175

actgacctag agaagctttt agcacctggt tggagggggt ttatagcaac cgcaaccctg	60
taaacctggt ttttgatga agctgaagct ataaaccccc tccaaccagg tgctaatttg	120
caaaagccta ggccctccaaa	140

<210> 176

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 176
agtcagtcga gagctcatgg atcccggtc tatgctgatt ggctatgtaa gg 52

<210> 177
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 177
atgctgatat ctagactatc ccgatttagg tgggcgacc 39

<210> 178
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 178
agtcagtcga gagctcatgc tgattggcta tgtaagg 37

<210> 179
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 179
tctagactac ggatccaccg atttacgcgg gc 32

<210> 180
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 180
atcgcgtatc tagactagcc gaggcaggcc aaggcgacg 39

<210> 181
<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 181

atcgcgatc tagactagct catctgaac tgcgtcatg 39

<210> 182

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 182

gtcgcccgcg taaatcgga tccactgcag atcggggggg ggc 43

<210> 183

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 183

gtcgcccgcg taaatcgga tccccctgc ctgcgttctc ggc 43

<210> 184

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 184

gtcgcccgcg taaatcgga tccgattcga tgcctgccgt cgg 43

<210> 185

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 185
gtcgcccgcg taaatcggga tccaccgtgc gtgtcgctgt cactg 45

<210> 186
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 186
gtcgcccgcg taaatcggga tccgtggatc tacgcacgct cggc 44

<210> 187
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 187
gtcgcccgcg taaatcggga tccacacacg cgcacatcgt tgc 43

<210> 188
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 188
gtcgcccgcg taaatcggga tcccacgaag acatcggttg cgtcg 45

<210> 189
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 189
gtcgcccgcg taaatcggga tccagcgctc tggaggcctt gctc 44

<210> 190
<211> 44
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 190

gtcgcccgcg taaatcggga tccttgaca caggccaact tctc 44

<210> 191

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 191

gtcgcccgcg taaatcggga tccagcgcg tgaccgcagt gga 43

<210> 192

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 192

ggatccgat ttacgcggc 20

<210> 193

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 193

atcgtagcag ctacgccac catgctgatt ggctatgtaa g 41

<210> 194

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 194

ggatccagac cccgatttac gcgggc 26

<210> 195
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 195
ttagcacctg gttggagggg gtttatatcc aaaaccatgg ttacagtat aaacccccctc 60
caaccaggtg ctaa 74

<210> 196
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 196
ttagcacctg gttggagggg gtttataagg ttttggtacc aaatgtctat aaacccccctc 60
caaccaggtg ctaa 74

<210> 197
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 197
tataaaacccc ctccaaccag gtgctaactg taaaccatgg ttttggtatta gcacctgggt 60
ggaggggggtt tata 74

<210> 198
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 198
ttagcacctg gttggagggg gtttataaaa accatgggtt atataaaccc cctccaacca 60
ggtgctaa 68

<210> 199

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 199

ttagcacctg gttggagggg gtttatagct tccaaaacca tggtttacag ggttataaac 60

cccctccaac caggtgctaa 80

<210> 200

<211> 86

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 200

ttagcacctg gttggagggg gtttatagct tcatacaaaa ccatgggtta cagggttcct 60

ataaaccccc tccaaccagg tgctaa 86

<210> 201

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 201

ttagcacctg gttggagggg gtttatagct tcagcttcat ccaaaaccat ggtttacagg 60

gttcgggttc ctataaaccc cctccaacca ggtgctaa 98

<210> 202

<211> 74

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 202

ttagcacctg gttggagggg gtttatatcc aaaaccgggg ttacagtat aaacccctc 60

caaccaggtg ctaa 74

<210> 203

<211> 74

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 203
ttagcacctg gttggagggg gtttatacga aatattataa attatcatat aaacccccctc 60
caaccagggtg ctaa 74

<210> 204
<211> 86
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 204
ttagcacctg gttggagggg gtttatagct tcacccaaaa ccgggggttta cagggttcct 60
ataaaccccc tccaaccagg tgctaa 86

<210> 205
<211> 86
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 205
ttagcacctg gttggagggg gtttatagct tcacgaaata ttataaatta tcagggttcct 60
ataaaccccc tccaaccagg tgctaa 86

<210> 206
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 206
gcgggaggcg tgtccaaaac catgggtttac aggggttccta taaacccccct ccaaccagggt 60
gctaa 65

<210> 207
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 207
gtggtgtaca gtagggggag atgcatccaa aaccatgggt tacagtgcac ctccccctac 60
tgtacaccac 70

<210> 208
<211> 82
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 208
gtggtgtaca gtagggggag atgcagctgc ttccaaaacc atgggtttaca ggtgggttgc 60
atctccccct actgtacacc ac 82

<210> 209
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 209
Gln Trp Ser Gly
1

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

<220>
<223> Synthetic construct

<400> 210
Arg Ser Asn Gly
1

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

<220>
<223> Synthetic construct

<400> 211

Ser Arg Ser Gly
1

<210> 212

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 212

Gln Trp Ser Gly
1

<210> 213

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 213

gataaacccc ctcaa 16

<210> 214

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 214

ataaaccccc tcaa 15

<210> 215

<211> 35

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 215

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

Pro Leu Gln
35

<210> 216
<211> 35
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 216
Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala
1 5 10 15

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

Pro Leu Gln
35

<210> 217
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 217
Lys Gln Trp Ser Gly
1 5

<210> 218
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 218
Lys Arg Ser Asn Gly
1 5

<210> 219
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 219

Lys Ser Arg Ser Gly
1 5

<210> 220

<211> 35

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 220

Ala Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala
1 5 10 15

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

Pro Leu Gln
35

<210> 221

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 221

Lys Gln Trp Ser Gly
1 5

<210> 222

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 222

aataaacccc ctcaa 16

<210> 223

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 223
tataaacccc ctcaa 16

<210> 224
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 224
Lys Arg Gly Gly
1

<210> 225
<211> 34
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 225
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys
1 5 10 15

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

His Gly

<210> 226
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 226
Ser His Asp Gly
1

<210> 227
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 227
Ala Ser His Asp Gly Gly
1 5

<210> 228
<211> 81
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> misc_feature
<222> (50)..(51)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (53)..(54)
<223> n is a, c, g, or t

<400> 228
tctcaactcc ccgcctccg tgagcaaggc ctccagagcg cgtgccccmn nnnntttgcc 60
gacgccaacg atgtcttcgt g 81

<210> 229
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> misc_feature
<222> (47)..(48)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (50)..(51)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (53)..(54)
<223> n is a, c, g, or t

<220>

<221> misc_feature
<222> (56)..(57)
<223> n is a, c, g, or t

<400> 229
tctcaactcc cccgcctccg tgagcaaggc ctccagagcg cgtgcmnnmn nmnnmnnnttt 60
gccgacgcca acgatgtctt cgtg 84

<210> 230
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> misc_feature
<222> (43)..(44)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (46)..(47)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (49)..(50)
<223> n is a, c, g, or t

<400> 230
cccgcctccg tgagcaaggc ctccagggcg cgtgcgccgg amnnmnnmnn gccgacgcca 60
acgatgtctt cgtgtgtcgc 80

<210> 231
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<220>
<221> misc_feature
<222> (53)..(54)
<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (56)..(57)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (59)..(60)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (62)..(63)

<223> n is a, c, g, or t

<400> 231

ggcaccgcgtc agtgcattgc gccatgcatg cactgcctcc actgcggtca cmnnmnnmnn 60

mnntgcaatc ttgagaagtt ggcctgtgtc 90

<210> 232

<211> 71

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 232

agagagagaa gaaaatgaga tctcctaaga aaaagaggaa ggtgcaggtg gatctacgca 60

cgctcggcta c 71

<210> 233

<211> 71

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 233

aggaagaaga gaagcatgag atctcctaag aaaaagagga aggtgatggt ggacttgagg 60

acactcgggtt a 71

<210> 234

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 234
aagagaagaa gaagaagcat tgcgcatgc atgcactgcc tcta 44

<210> 235
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 235
cccgccaccc accgtgc 17

<210> 236
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 236
tgctctatgc atgcactgcc tcc 23

<210> 237
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 237
agagaagaga agagaaggcg ccgcgcccc aggcggcctc gggatcccct cggcctccgc 60
gcgccaag 68

<210> 238
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 238
agagagagag agagagtcta gaggccggcc tggccgctca tcccgaactg cgtcatggcc 60
tcata 65

<210> 239

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 239

gcccagatc ctggtacgct ctagagg 27

<210> 240

$\langle 211 \rangle$ 70

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

$\langle 400 \rangle$ 240

atcttagcac ctggttgag ggggtttatt gggttttccc aataaacccc ctccaaccag 60

gtgctaagat 70

<210> 241

<211> 70

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 241

atcttagcac ctggttggag ggggtttata ggggtttccc tataaacccc ctccaaccag 60

gtgctaagat 70

$\langle 210 \rangle$ 242

$\langle 211 \rangle$ 70

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 242

atcttagcac ctggttggag ggggtttatc ggggtttccc gataaacccc ctccaaccag 60

gtgctaagat 70

<210> 243

$\langle 211 \rangle$ 70

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 243

atcttagcac ctggttgag ggggttatg gggtttccc cataaacccc ctccaaccag	60
gtgctaagat	70

<210> 244

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 244

ttaaaagcca ggacggtcac	20
-----------------------	----

<210> 245

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 245

tgtagggagc ccagaagaga	20
-----------------------	----

<210> 246

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 246

acagtttgca ttcattgagg gc	22
--------------------------	----

<210> 247

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 247

ccgagcgcgc aagctcagtt 20

<210> 248

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 248

cgcgcatccc cgcccagtg gactttg 27

<210> 249

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 249

ccggaattca cctgtagag ctactgc 27

<210> 250

<211> 54

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 250

agagagagag agaggcggcc gccctaccag ggatttcagt cgatgtacac gttc 54

<210> 251

<211> 58

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 251

aagaagaaga aggaagagaa gtaggcctgt catcgtcggg aagacctgcg acacctgc 58

<210> 252

<211> 138

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 252

actgctatcc gagtataaac cccctccaac caggtataaa ccccctccaa ccaggtataa	60
accccctcca accaggtata aaccccctcc aaccaggtat aaaccccctc caaccaggat	120
ctgcgatcta agtaagct	138

<210> 253

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 253

ttaattaaga gtctagatta gcacctggtt ggaggggggtt tattgcttca tccaaaacca	60
tggtttacag ggttccaata aaccccctcc aaccaggtgc taaagatcta ggaggaattt	120
aaaatgag	128

<210> 254

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 254

actgacctag agaagctttt agcacctggt tggaggggggt ttattgcaac cctgtaaacc	60
atggtttttg atgaagcaat aaaccccctc caaccagggtg ctaactgcag ttatttgtac	120
agttcatc	128

<210> 255

<211> 128

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 255

ttaattaaga gtctagatta gcacctggtt ggaggggggtt tatcgcttca tccaaaacca	60
tggtttacag ggttccgata aaccccctcc aaccaggtgc taaagatcta ggaggaattt	120
aaaatgag	128

<210> 256
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 256
actgacctag agaagctttt agcacctggg tggagggggg ttatcgcaac cctgtaaacc 60
atggtttttg atgaagcgat aaacccctc caaccagggt ctaactgcag ttatttgtac 120
agttcatc 128

<210> 257
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 257
ttaattaaga gtctagatta gcacctgggt ggaggggggt tatggcttca tccaaaacca 60
tggtttacag gggtcccata aacccctcc aaccagggtc taaagatcta ggaggaattt 120
aaaatgag 128

<210> 258
<211> 128
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 258
actgacctag agaagctttt agcacctggg tggagggggg ttatggcaac cctgtaaacc 60
atggtttttg atgaagccat aaacccctc caaccagggt ctaactgcag ttatttgtac 120
agttcatc 128

<210> 259
<211> 144
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 259
actgctatct cgagctataa accccctcca accaggctat aaacccctc caaccagggt 60

ataaaccccc tccaaccagg ctataaaccc cctccaacca ggctataaac cccctccaac 120
caggatctgc gatctaagta agct 144

<210> 260

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 260

actgctatct cgagcaataa accccctcca accaggcaat aaaccacctc caaccaggca 60
ataaaccccc tccaaccagg caataaaccc cctccaacca ggcaataaac cccctccaac 120
caggatctgc gatctaagta agct 144

<210> 261

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 261

actgctatct cgagccataa accccctcca accaggccat aaaccacctc caaccaggcc 60
ataaaccccc tccaaccagg ccataaaccc cctccaacca ggccataaac cccctccaac 120
caggatctgc gatctaagta agct 144

<210> 262

<211> 144

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 262

actgctatct cgagcgataa accccctcca accaggcgat aaaccacctc caaccaggcg 60
ataaaccccc tccaaccagg cgataaaccc cctccaacca ggcgataaac cccctccaac 120
caggatctgc gatctaagta agct 144

<210> 263

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 263

tcagaaacag ctcttcttca aatct 25

<210> 264

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 264

ttaattaaga gtctagagga ggcgtgtcca aaaccatggt ttacagcacg cctccagatc 60

taggaggaat ttaaaatgag 80

<210> 265

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 265

actgacctag agaagcttgg aggcgtgctg taaaccatgg ttttggacac gcctccctgc 60

agttatttgc acagttcatc 80

<210> 266

<211> 86

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 266

ttaattaaga gagatctgct gatgcagata cagaaaccaa ggTTTTctta cttgctgctg 60

cgcgatctgc atctcaatta gtcagc 86

<210> 267

<211> 82

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 267

caggaccagg gatccgcagg agcaagtaag aaaacgttgg tttctgtatc tgcacagca 60

caccacccacg gaccgcgagc agcagcgcag aaaaaccccg ccccgacac cgcacccagc
 atttcgataa gccagtaagc ag

82

<210> 268
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 268
caccaccacg cgcgcaagct tagatctggc ccaggcggcc accatgctga ttggctatgt 60
aaggg 65

<210> 269
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 269
caccaccaca ccggttcccg atttaggtgg gcgac 35

<210> 270
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 270
gttcctgccaggatccacta g 21

<210> 271
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 271
gcatgtgtcc agatgcatag g 21

$\langle 210 \rangle$ 272
 $\langle 211 \rangle$ 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 272

caccttctcc caggataagg 20

<210> 273

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 273

gttggcctgt attcctctgg 20

<210> 274

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 274

aatgaagttc ccttggcact tc 22

<210> 275

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 275

ctgaagggtt ttaagtcag aag 23

<210> 276

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 276
tgacgtcaat gacggtaa at gg 22

<210> 277
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

```
<400> 277
actgacctag agaagctttt agcacctggt tggaggggggt ttataaccct gtaaaccatg      60
gttttggaag ctataaaccc cctccaacca ggtgctaact gcagttattt gtacagttca      120
tc                                          122
```

<210> 278
<211> 140
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

```
<400> 278
actgacctag agaagctttt agcacctggg tggagggggg ttatagcaac cgcaaccctg      60
taaaccatgg ttttggatga agctgaagct ataaaccccc tccaaccagg tgctaactgc      120
agttattttg acagttcatc                                     140
```

<210> 279
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 279
gcgggaggcg tg 12

<210> 280
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 280

tcttcattac acctgca 17

<210> 281

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 281

cagtcagtat caatt 15

<210> 282

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 282

cctgcagctc tcatttt 17

<210> 283

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 283

attcttccag aattga 16

<210> 284

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 284

cagaattgat actgact 17

<210> 285

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 285

tcattacacc tgcagc 16

<210> 286

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 286

cttccagaat tgatactg 18

<210> 287

$\langle 211 \rangle$ 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 287

ataaaccccc tccaa 15

<210> 288

<211> 74

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 288

ttagcacctg gttggagggg gtttatatcc aaaaccatgg ttacagtat aaaccccctc 60

caaccaggtg ctaa 74

<210> 289

<211> 74

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 289

[illegible]

ttagcacctg gttggagggg gtttataagg ttttggtagc aaatgtctat aaacccccctc 60
caaccaggtg ctaa 74

<210> 290
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 290
tataaaccct ctccaaccag gtgctaactg taaaccatgg ttttggatta gcacctgggt 60
ggaggggggt tata 74

<210> 291
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<220>
<221> misc_feature
<222> (2)..(3)
<223> Xaa can be any naturally occurring amino acid

<400> 291
Lys Xaa Xaa Gly Ala Arg
1 5

<210> 292
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 292
tggaaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga 60
agaccttctt tttgagatct ggt 83

<210> 293
<211> 77
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 293
tggaaattct tccataattg atattgactg tatggaaggc tgcgggtgta atgaatacct 60
tcttttttgag atctggt 77

<210> 294
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 294
tggaaattct tccagaattg atactgactg tatggaaaac tgcaggtgta atgaagacct 60
tcttttttgag atctggt 77

<210> 295
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 295
tggaaattct tccagaattg atactgactg tatggaaagc tgcaggtgta atgaagacct 60
tcttttttgag atctggt 77

<210> 296
<211> 76
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 296
tggaaattct tccagaattg atactgactg tatggaagct gcaggtgtaa tgaagacctt 60
cttttttgaga tctggt 76

<210> 297
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 297	
tggaattct tccagaattg atactgactg tagagctgca ggtgtaatga agaccttctt	60
tttgagatct ggt	73
<210> 298	
<211> 72	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 298	
tggaattct tccagaattg atactgactg tatgctgcag gtgtaatgaa gaccttcttt	60
ttgagatctg gt	72
<210> 299	
<211> 70	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 299	
tggaattct tccagaattg atactgactg tgctgcaggt gtaatgaaga ccttcttttt	60
gagatctggt	70
<210> 300	
<211> 83	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 300	
tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
agaccttctt tttgagatct ggt	83
<210> 301	
<211> 79	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 301	
tggaattct tccagaattg atactgactg tatggaaaga gctgcaggtg taatgaagac	60

cttcttttttg agatctggt	79
<210> 302	
<211> 78	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 302	
tggaaattct tccagaattg atactgactg tatggaagag ctgcaggtgt aatgaagacc	60
ttcttttttga gatctggt	78
<210> 303	
<211> 77	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 303	
tggaaattct tccagaattg atactgactg tatggagagc tgcaggtgta atgaagacct	60
tcttttttgag atctggt	77
<210> 304	
<211> 74	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 304	
tggaaattct tccagaattg atactgacta tgagagctgc aggtgtaatg aagaccttct	60
ttttgagatc tgggt	74
<210> 305	
<211> 73	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 305	
tggaaattct tccagaattg atactgactg tagagctgca ggtgtaatga agaccttctt	60
tttgagatct ggt	73

<210> 306
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 306
tggaaattct tccagaattg atactgactg agagctgcag gtgtaatgaa gaccttcttt 60

ttgagatctg gt 72

<210> 307
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 307
tggaaattct tccaggattg atactgactg agagctgcag gtgtaatgaa gaccttcttt 60

ttgagatctg gt 72

<210> 308
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 308
tggaaattct tccagaattg atactgactg agagctgcag gtgtaatgaa gaccttcttt 60

ttgagatctg gt 72

<210> 309
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 309
tggaaattct tccagaattg atactgacta ggagctgcag gtgtaatgaa gaccttcttt 60

ttgagatctg gt 72

<210> 310

<211> 83

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 310

tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga 60

agaccttctt tttagatct ggt 83

<210> 311

<211> 81

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 311

tggaattct tccagaattg atactgacta tggaaaatga gagctgcagg tgtaatgaag 60

accttctttt tgagatctgg t 81

<210> 312

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 312

tggaattct tcctcaattg atactgatat ggaaaatgag agctgcagg gtaatgaaga 60

ccttcttttt gagatctgg 80

<210> 313

<211> 74

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 313

tggaattct tccagaattg atatggaaaa tgagagctgc aggtgtaatg aagaccttct 60

ttttgagatc tgg 74

<210> 314

<211> 83

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 314

tggaaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
agaccttctt tttgagatct ggt	83

<210> 315

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 315

tggaaattct tccagaattg atactgactg tatgatgaga gctgcaggtg taatgaagac	60
cttcttttttg agatctggt	79

<210> 316

<211> 79

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 316

tggaaattct tccagaattg atactgactg tatggtgaga gctgcaggtg taatgaagac	60
cttcttttttg agatctggt	79

<210> 317

<211> 78

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 317

tggaaattct tccagaattg atactgactg taaatgagag ctgcaggtgt aatgaagacc	60
ttcttttttga gatctggt	78

<210> 318

<211> 78

<212> DNA

<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 318
tggaaattct tccagaattg atactgactg tatgtgagag ctgcaggtgt aatgaagacc 60
ttctttttga gatctgggt 78

<210> 319
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 319
tggaaattct tccagaattg atactgaaaa tgagagctgc aggtgtaatg aagaccttct 60
ttttgagatc tggt 74

<210> 320
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 320
tggaaattct tccagaattg atactgactg tatggactgc aggtgtaatg aagaccttct 60
ttttgagatc tggt 74

<210> 321
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 321
tggaaattct tccagaattg atactgactg tgagagctgc aggtgtaatg aagaccttct 60
ttttgagatc tggt 74

<210> 322
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 322	
tggaattct tccagaattg atactggtat gagagctgca ggtgtaatga agaccttctt	60
tttgagatct ggt	73
<210> 323	
<211> 71	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 323	
tggaattct tccagaattg atactgactg tagctgcagg tgtaatgaag accttctttt	60
tgagatctgg t	71
<210> 324	
<211> 83	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 324	
tggaattct tccagaattg atactgactg tatggaaaat gagagctgca ggtgtaatga	60
agaccttctt tttgagatct ggt	83
<210> 325	
<211> 75	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 325	
tggaattct tccagaattg atactgactg tatggaaact gcggtgtaat gaagaccttc	60
tttttgagat ctggt	75
<210> 326	
<211> 76	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Synthetic construct	
<400> 326	
.....	76

tggaaattct tccagaattg atactgactg tatgagagct gcaggtgtaa tgaagacctt 60
ctttttgaga tctggg 76

<210> 327
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 327
tggaaattct tccagaatta atactgactg tgagagctgc aggtgtaatg aagaccttct 60
ttttgagatc tggg 74

<210> 328
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 328
tggaaattct tccagaattg atactgactg gagagctgca ggtgtaatga agaccttctt 60
tttgagatct ggt 73

<210> 329
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 329
tggaaattct tccagaattg atactgactg agagctgcag gtgtaatgaa gaccttcttt 60
ttgagatctg gt 72

<210> 330
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 330
tggaaattct tccagaattg atactgactg ctgcaggtgt aatgaagacc ttctttttga 60
gatctggg 68

<210> 331
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 331
tggaaattct tccagaattg atactgactg tatgtgtaat gaagaccttc tttttgagat 60
ctggt 65

<210> 332
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 332
tggaaattct tccagaattg atacttttg agatctgt 39

<210> 333
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 333
Arg Asn Asn Arg Asn Asn Arg Asn Asn Arg Asn Asn Asn Asn Tyr Asn
1 5 10 15
Asn Tyr Asn Asn Tyr Asn Asn Tyr
20

<210> 334
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 334
Arg Asn Asn Arg Asn Asn Arg Asn Asn Arg Asn Asn Asn Asn Tyr Asn
1 5 10 15
Asn Tyr Asn Asn Tyr Asn Asn Tyr
20

--

<210> 335
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 335
tccaaaacca tggtttacag 20

<210> 336
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 336
tccaaaacca tggtttacag 20

<210> 337
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 337
tccaaaacca tggtttacag 20

<210> 338
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 338
tccaaaacca tggtttacag 20

<210> 339
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 339

tccaaaacca tggtttacag 20

<210> 340

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (17)..(18)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (21)..(22)

<223> n is a, c, g, or t

<400> 340

ggaggcgtgt ccaaaannat nntttacagc acgcctcc 38

<210> 341

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 341

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15

Asp Ile Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30

Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Cys Asp Arg Lys Gly Leu

35

40

45

Asp Leu Leu Arg Met Lys Val Glu Glu Gly Asp Val Ile Leu Val Lys
50 55 60

Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Ile Gln Leu Ile
65 70 75 80

Lys Glu Phe Asp Ala Gln Gly Val Ser Ile Arg Phe Ile Asp Asp Gly
85 90 95

Ile Ser Thr Asp Gly Glu Met Gly Lys Met Val Val Thr Ile Leu Ser
100 105 110

Ala Val Ala Gln Ala Glu Arg Gln Arg Ile Leu Glu Arg Thr Asn Glu
115 120 125

Gly Arg Gln Glu Ala Met Ala Lys Gly Val Val Phe Gly Arg Lys Arg
130 135 140

<210> 342

<211> 141

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 342

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15

Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30

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

Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60

Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80

Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95

Ser Ser Pro Met Gly Arg Phe Phe Phe His Val Met Gly Ala Leu Ala
100 105 110

Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125

Ala Ala Arg Asn Lys Gly Arg Ile Gly Gly Arg Pro Pro
130 135 140

<210> 343

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>
<221> misc_feature
<222> (2)..(3)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (5)..(6)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (8)..(9)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (11)..(16)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (21)..(21)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (29)..(34)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (36)..(37)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (39)..(40)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (42)..(43)

<223> n is a, c, g, or t

<400> 343

rnnrrnnrrnr nnnnnnaaab nwwnvtttnn nnnnynnnynn ynny 44

<210> 344

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 344

gctgatgcag atacagaaac caaggtttc ttacttgctg ctgc 44

<210> 345

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 345

gtggatggag cagccaatag gttcctttcc tcccccttag cccc 44

<210> 346

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 346

agggaagtca atccagaaac catcctttat cccttcctgt cctt 44

<210> 347

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 347

ggaaatgtaa aagtagaaac taaagtttct gctttcattc ttcc 44

<210> 348

<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 348
ggaagaagga tgagagaaac taaccttgt ggaaccctg cagc 44

<210> 349
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 349
aacggcagaa gaagaaaaat tatactttct ttccattgt ttc 44

<210> 350
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 350
gaggtaaata cttgataaat gttgctttt tccccatta ccct 44

<210> 351
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 351
attgtggatg gagtaaaaat gacctttaa tacatttcta catt 44

<210> 352
<211> 44
<212> DNA
<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 352
ataggagaaa atttggaag tataat tttt cagactactc tttt 44

<210> 353
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 353
acagaagaca ttaagaaaac ctaacttgac ctctatggt tcc 43

<210> 354
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 354
ggcaggacag ctaactaatg aaagggttgg tgtgtgtctg tctt 44

<210> 355
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 355
agggatgagg cctcataaag taaagtttt tgttgtttg ttcc 44

<210> 356
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 356
acagtcaaag tatttgaaag ttaacttttt tcgtcagctc ttcc 44

<210> 357

<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 357
gaaattgtgg acaattaaat tatcctttct gggcccctta tttc 44

<210> 358
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 358
gaaattggaa ggaaaaaaat tatcctttat ggtgtaatac ttat 44

<210> 359
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 359
aaaacagctg gcttgaaag gaaactttta actactatcc tgcc 44

<210> 360
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 360
atagtaagtg ctcaataaat gttcgtttat atcatcattg tggc 44

<210> 361
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 361
aaagatggaa caaacaaaat taaggtttag tacattataa ttcc 44

<210> 362
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 362
gcgggaggcg tgtccaaacc atggtttaca gcacgcctcc cgc 43

<210> 363
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 363
gctgatgcag atcgagaaac caaggtttc ttacttgctg ctgc 44

<210> 364
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 364
gtggatggag cagccaatag gttcctttcc tcccccttag cccc 44

<210> 365
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic construct

<400> 365
agggaagtca atccagaaac catcctttat cccttctgt cctt 44

<210> 366
<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 366

agatacagaa accgttttct tacttgctgc tggcc 35

<210> 367

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<400> 367

tcagggaagt catcctttat cccttcctgt ccttagct 38

<210> 368

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (8)..(9)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (12)..(13)

<223> n is a, c, g, or t

<400> 368

tccasssna tnnssacag 20

<210> 369

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (1)..(4)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (8)..(9)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (12)..(13)

<223> n is a, c, g, or t

<220>

<221> misc_feature

<222> (17)..(20)

<223> n is a, c, g, or t

<400> 369

nnnnaaanna tnntttnnnn 20

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO2008006028A [0047]
- US7101972B [0047]
- US7067617B [0047]
- US6790941B [0047]
- US6610512B [0047]
- US6242568B [0047]
- US6140466A [0047]
- US6140081A [0047]
- US20060223757A [0047]
- US20060211846A [0047]

- [US20060078880A \[0047\]](#)
- [US20050148075A \[0047\]](#)
- [US20050084885A \[0047\]](#)
- [US20040224385A \[0047\]](#)
- [US20030059767A \[0047\]](#)
- [US20020165356A \[0047\]](#)
- [US7067317B \[0047\]](#)
- [US7030215B \[0047\]](#)
- [US7026462B \[0047\]](#)
- [US7013219B \[0047\]](#)
- [US6979539B \[0047\]](#)
- [US6933113B \[0047\]](#)
- [US6824978B \[0047\]](#)
- [US6794136B \[0047\]](#)
- [US6785613B \[0047\]](#)
- [US6777185B \[0047\]](#)
- [US6706470B \[0047\]](#)
- [US6607882B \[0047\]](#)
- [US6599692B \[0047\]](#)
- [US6534261B \[0047\]](#)
- [US6503717B \[0047\]](#)
- [US6453242B \[0047\]](#)
- [US20060246588A \[0047\]](#)
- [US20060246567A \[0047\]](#)
- [US20060166263A \[0047\]](#)
- [US20060078878A \[0047\]](#)
- [US20050257062A \[0047\]](#)
- [US20050215502A \[0047\]](#)
- [US20050130304A \[0047\]](#)
- [US20040203064A \[0047\]](#)
- [US20030166141A \[0047\]](#)
- [US20030134318A \[0047\]](#)
- [US20030105593A \[0047\]](#)
- [US20030087817A \[0047\]](#)
- [US20030021776A \[0047\]](#)
- [US20020081614A \[0047\]](#)
- [US61818364B \[0189\]](#)
- [US61753763B \[0189\]](#)
- [US61696689B \[0189\]](#)

Non-patent literature cited in the description

- **MUSSOLINO et al.**A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicityNucleic Acids Research, 2011, vol. 39, 219283-9293 [0008]
- **MILLER et al.**A TALE nuclease architecture for efficient genome editing supplementary informationNature Biotechnology, 2010, vol. 29, 2143-148 [0008]
- **ESPOSITO, D.SCOCCA, J. J.**Nucleic Acids Research, 1997, vol. 25, 3605-3614 [0044]
- **NUNES-DUBY, S. E. et al.**Nucleic Acids Research, 1998, vol. 26, 391-406 [0044]
- **STARK, W. M. et al.**Trends in Genetics, 1992, vol. 8, 432-439 [0044]
- **WATSON et al.**Molecular Biology of the GeneBenjamin/Cummings19870000224- [0049]
- **CREIGHTON**ProteinsW. H. Freeman and Company19840000 [0049]
- **SCHULZSCHIMER**Principles of Protein StructureSpringer-Verlag19790000 [0049]
- Somat. Cell. Mol. Genet., 1981, vol. 7, 603-616 [0055]
- **SMITH et al.**Gene, 1988, vol. 67, 3140- [0058]
- **AUSUBEL et al.**Short Protocols in Molecular BiologyWiley & Sons19950000 [0070]
- **SMITHWATERMAN**Advances in Applied Mathematics, 1981, vol. 2, 482-489 [0073]
- **DAYHOFF**Atlas of Protein Sequences and StructureNational Biomedical Research Foundationvol. 3, 353-358 [0073]
- **GRIBSKOV**Nucl. Acids Res., 1986, vol. 14, 66745-6763 [0073]
- Wisconsin Sequence Analysis Package Program ManualGenetics Computer Group19950000 [0073]
- **SAMBROOK et al.**Molecular Cloning: A Laboratory ManualCold Spring Harbor19890000 [0076] [0078]
- Nucleic Acid Hybridization: A Practical ApproachIRL Press19850000 [0077]

PATENTKRAV

1. Isoleret polypeptid, der omfatter et *Xanthomonas*-afledt transskriptionsaktivator-lignende effektor (TALE)-protein med et N-terminalt domæne (NTD) omfattende en aminosyresekvens i N₁-hårnålen ifølge SEQ ID NO: 3 (VGKQWSGARAL) med én eller flere mutationer udvalgt fra: Q er Y, Q er S, Q er R, W er R, W er G, W er deleteret, S er R, S er H, S er A, S er N, og S er T.
5
2. Polypeptid ifølge krav 1, hvor NTD'et omfatter en aminosyresekvens udvalgt fra: VGKYRGARAL (SEQ ID NO: 4), VGKSRSGARAL (SEQ ID NO: 5), VGKYHGARAL (SEQ ID NO: 6) og VGKRGAGARAL (SEQ ID NO: 7).
10
3. Isoleret polypeptid, der omfatter et *Ralstonia*-afledt transskriptionsaktivator-lignende effektor (TALE)-protein med et N-terminalt domæne (NTD) omfattende en aminosyresekvens i N₁-hårnålen ifølge SEQ ID NO: 8 (IVDIAR₁QR₂SGDLA) med én eller flere mutationer udvalgt fra: R₁ er K, Q er Y, Q er S, Q er R, R₂ er W, R₂ er G, R₂ er deleteret, S er R, S er H, S er A, S er N, og S er T.
15
4. Polypeptid ifølge krav 3, hvor NTD'et omfatter en aminosyresekvens udvalgt fra: IVDIARQWSGDLA (SEQ ID NO: 9), IVDIARYRGDLA (SEQ ID NO: 10), IVDIARSRSGDLA (SEQ ID NO: 11), IVDIARYHGDLA (SEQ ID NO: 12) og IVDIARRGAGDLA (SEQ ID NO: 13).
20
5. Polypeptid ifølge et hvilket som helst af kravene 1-4, der endvidere omfatter et rekombinasedomæne eller et nukleasedomæne.
25
6. Polypeptid ifølge et hvilket som helst foregående krav, hvor TALE-proteinet omfatter en C-terminal eller N-terminal trunkering.
7. Polypeptid ifølge krav 1 eller 2, hvor TALE-proteinet er afledt af gruppen bestående af: AvrXa7, Tal1c og PthXo1.
30
8. Polypeptid ifølge krav 1 eller 2, hvor TALE-proteinet endvidere omfatter et gentaget variabelt di-rest (RVD)-domæne.

9. Polypeptid ifølge krav 8, hvor TALE-proteinet omfatter en C-terminal trunkering mellem aminosyreresterne 27 og 268, 92 og 134, 120 og 129, 74 og 147, eller 87 og 120 ifølge SEQ ID NO: 2.

10. Polypeptid ifølge krav 9, hvor TALE-proteinet er trunckeret ved aminosyreresten 28, 74, 87, 92, 95, 120, 124, 128, 129, 147 og 150 ifølge SEQ ID NO: 2.

11. Polypeptid ifølge et hvilket som helst af kravene 5-10, hvor rekombinasen er udvalgt fra gruppen af rekombinase katalytiske domæner bestående af Gin, Hin, Tn3, Sin, Beta, Pin, Min, Din, og Cin og muteiner af Gin, muteiner af Hin, muteiner af Sin, muteiner af Beta, muteiner af Pin, muteiner af Min, muteiner af Din, muteiner af Cin og muteiner af Tn3.

12. Isoleret nukleinsyremolekyle, der koder for polypeptidet ifølge et hvilket som helst af kravene 1-11.

13. Vektor, der omfatter nukleinsyremolekylet ifølge krav 12.

14. Værtscelle transformeret eller transficeret med nukleinsyremolekylet ifølge krav 12 eller vektoren ifølge krav 13.

15. Fremgangsmåde til generering af en transskriptionsaktivator-lignende effektor (TALE)-protein N-terminalt domæne (NTD), omfattende:

a) randomisering af en aminosyresekvens af NTD'et ved mutation af én eller flere aminosyrerester inde i NTD'et, hvor aminosyresekvensen er SEQ ID NO: 14 (VGKXXXGAR) eller SEQ ID NO: 15 (VDIAXXXXGDLA); og

b) udvælgelse af det randomiserede TALE-protein NTD fra (a), hvor TALE-proteinet NTD specifikt binder til et ønsket nukleotid eller har øget aktivitet.

16. Fremgangsmåde til stedsspecifik rekombination omfattende:

(a) tilvejebringelse af en DNA-sekvens, der omfatter mindst to bindingssteder for specifik interaktion med polypeptidet ifølge et hvilket som helst af kravene 1-11; og

(b) omsætning af DNA-sekvensen med polypeptidet, hvor polypeptidet katalyserer en stedsspecifik rekombinationshændelse, hvori begge strenge af DNA-sekvensen spaltes mellem de to steder, der specifikt interagerer med polypeptidet.

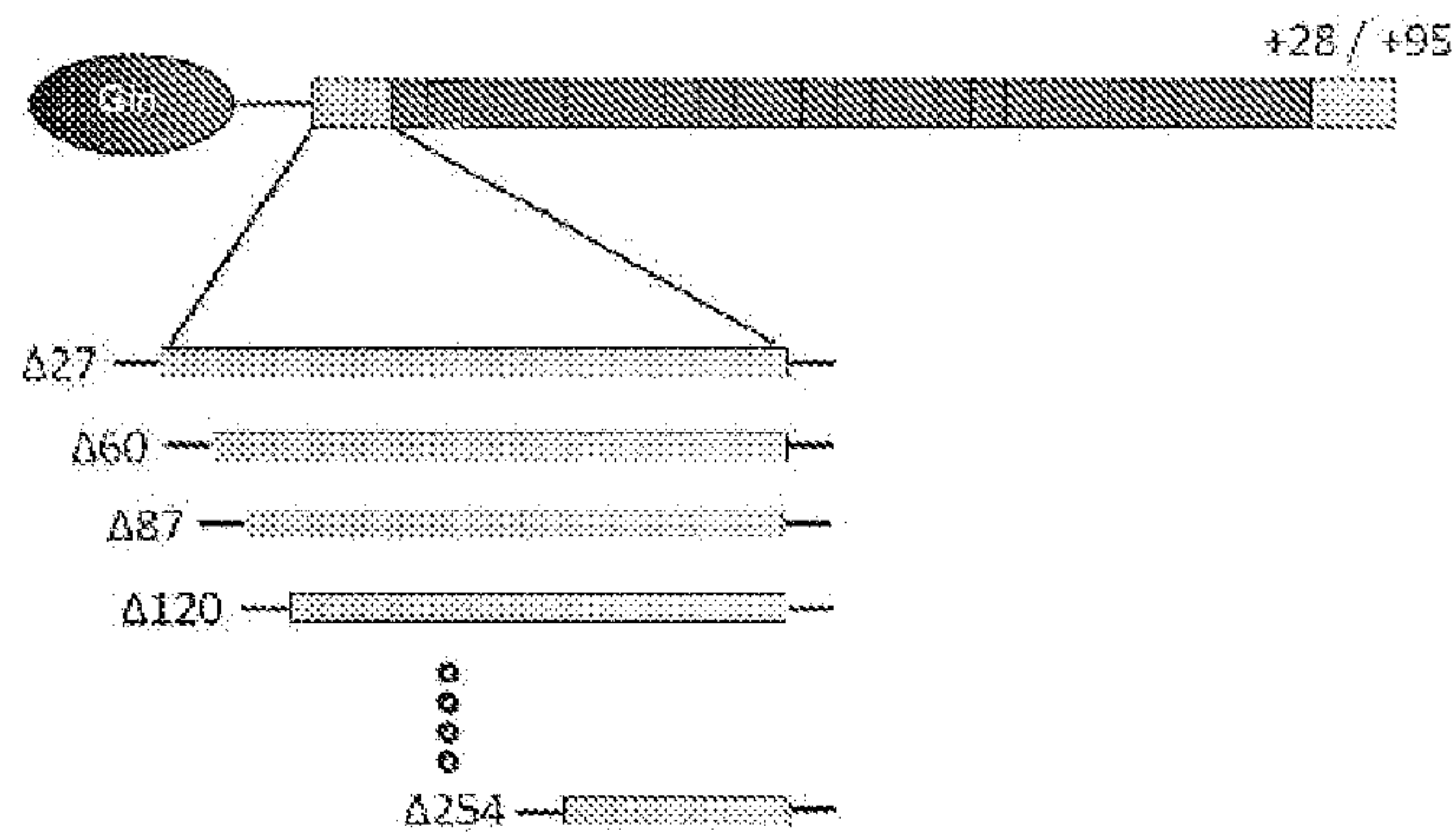
17. Sammensætning, der omfatter et nukleinsyremolekyle, der koder for polypeptidet ifølge et hvilket som helst af kravene 1-11, til anvendelse i genterapi.

5 **18.** Farmaceutisk sammensætning, der omfatter:

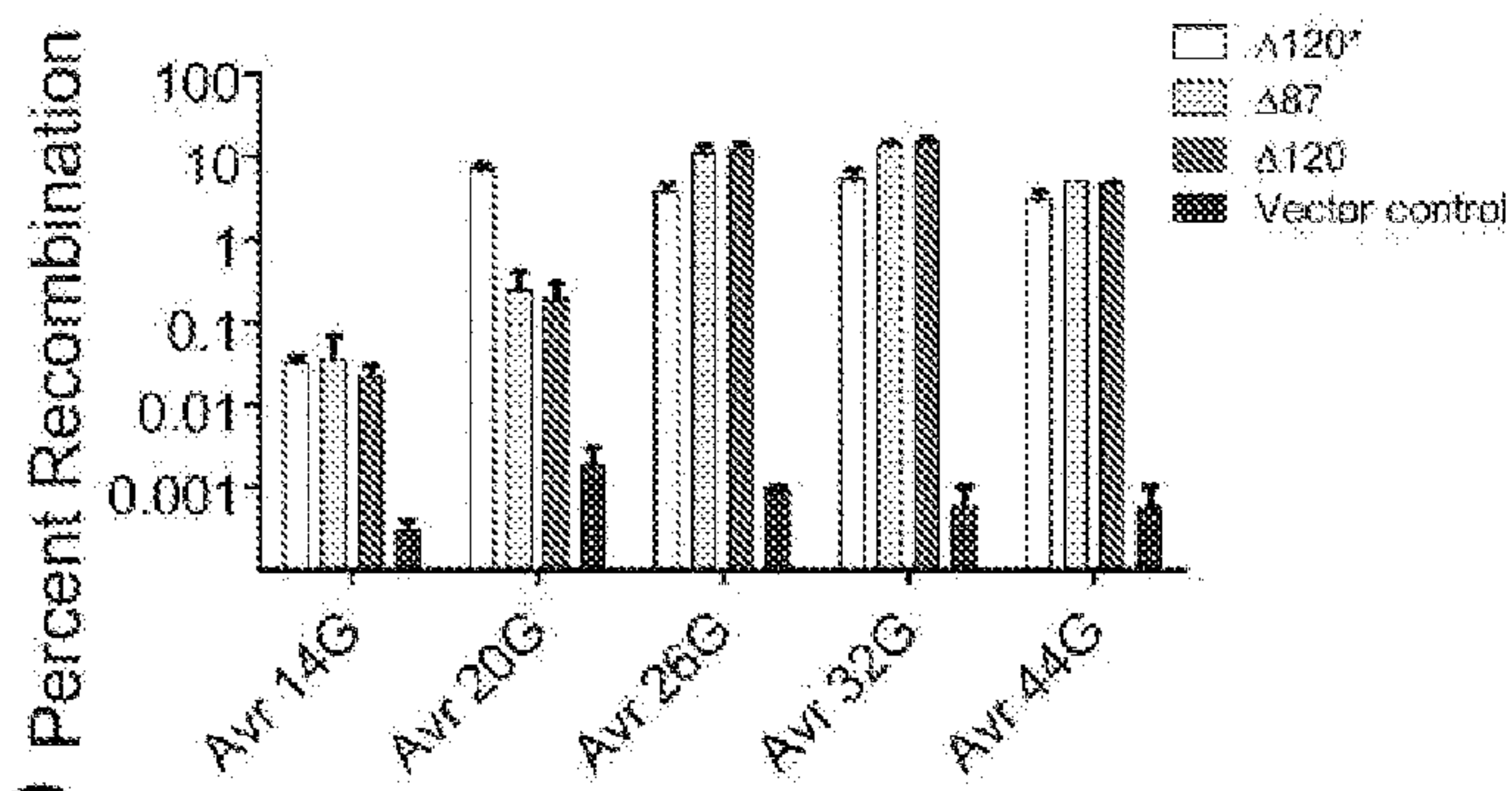
a) polypeptidet ifølge et hvilket som helst af kravene 1-11, eller en nukleinsyre, der koder for polypeptidet; og

b) en farmaceutisk acceptabel bærer.

A



B



C

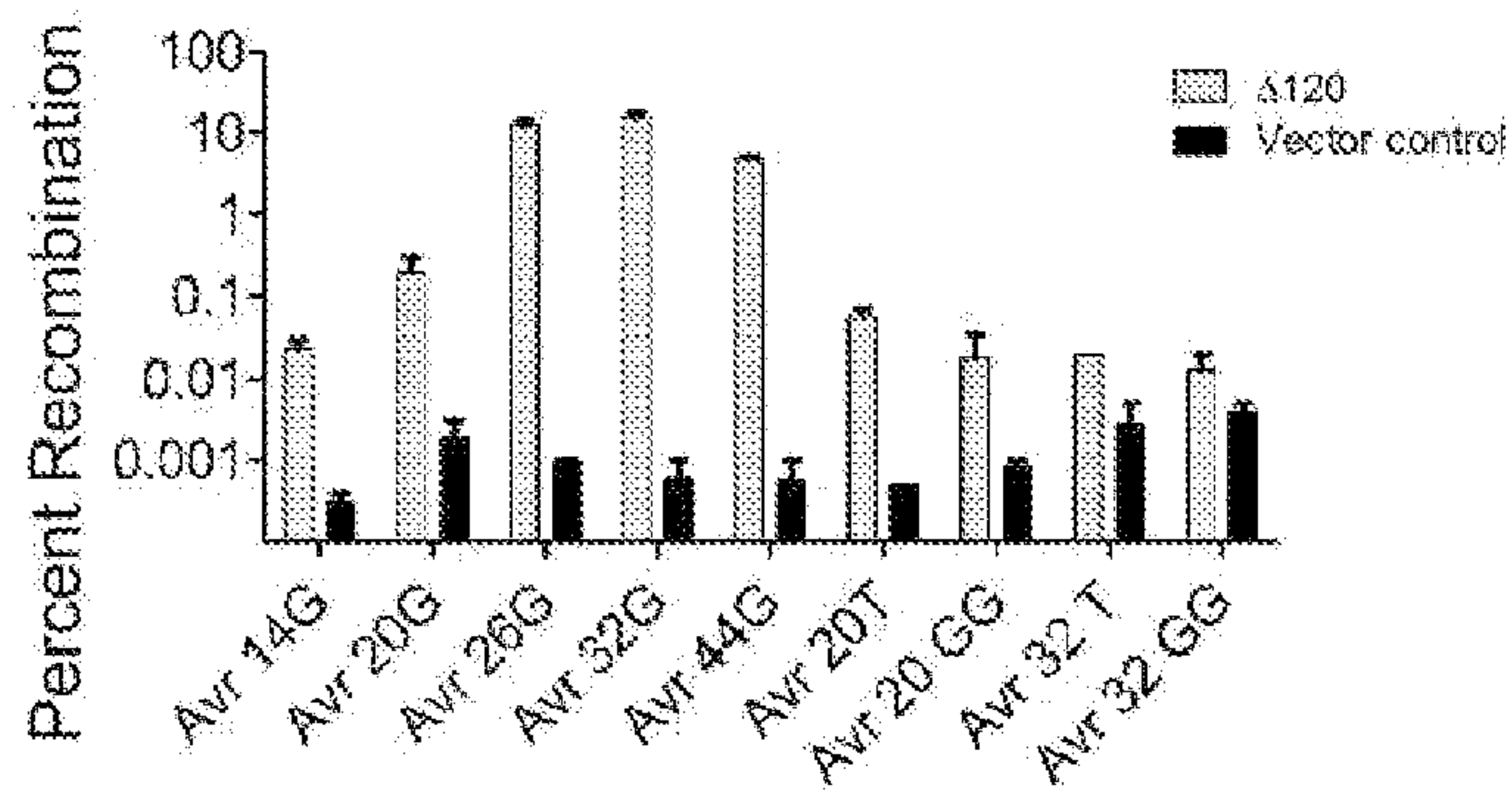


FIG. 2

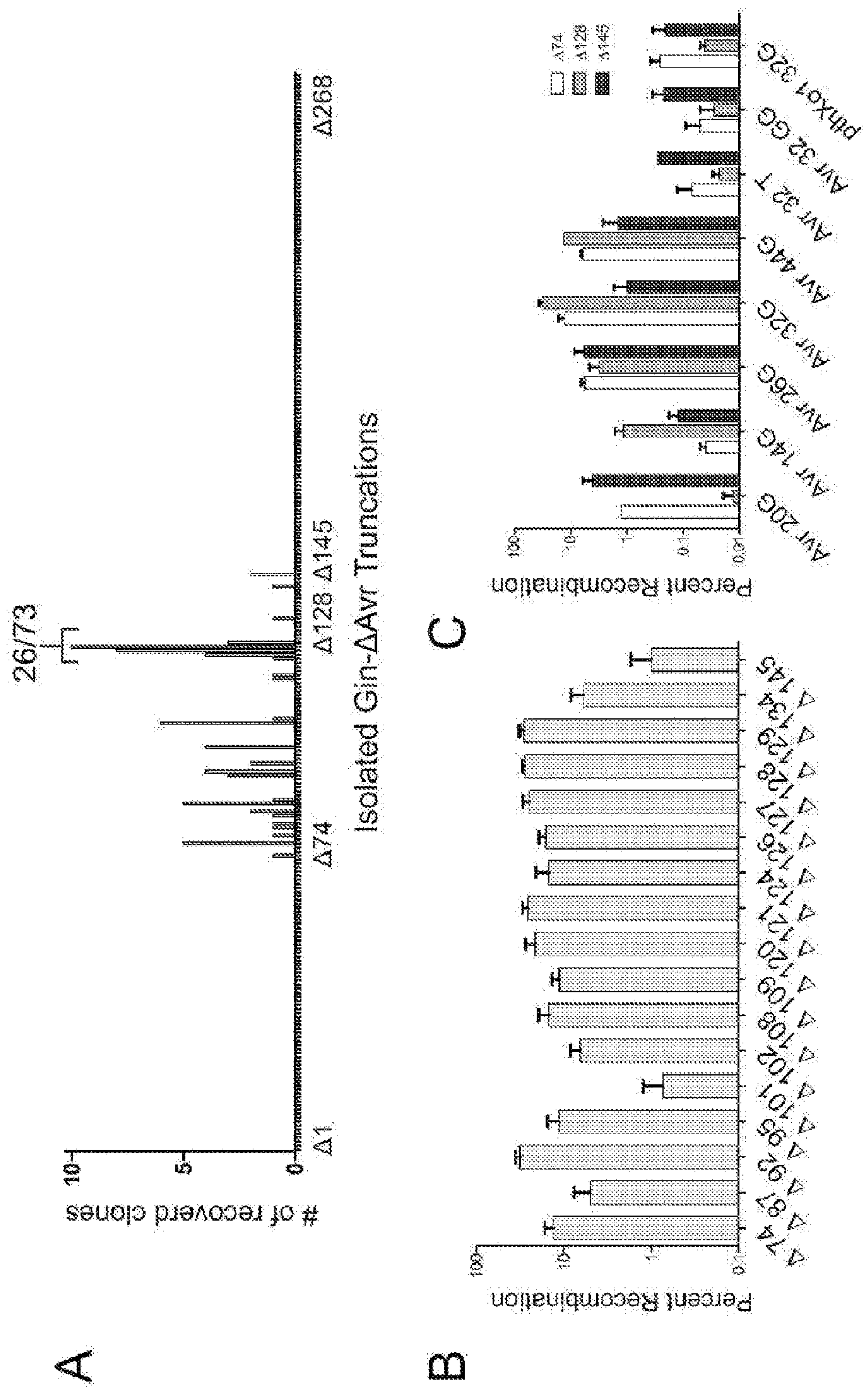


FIG. 3

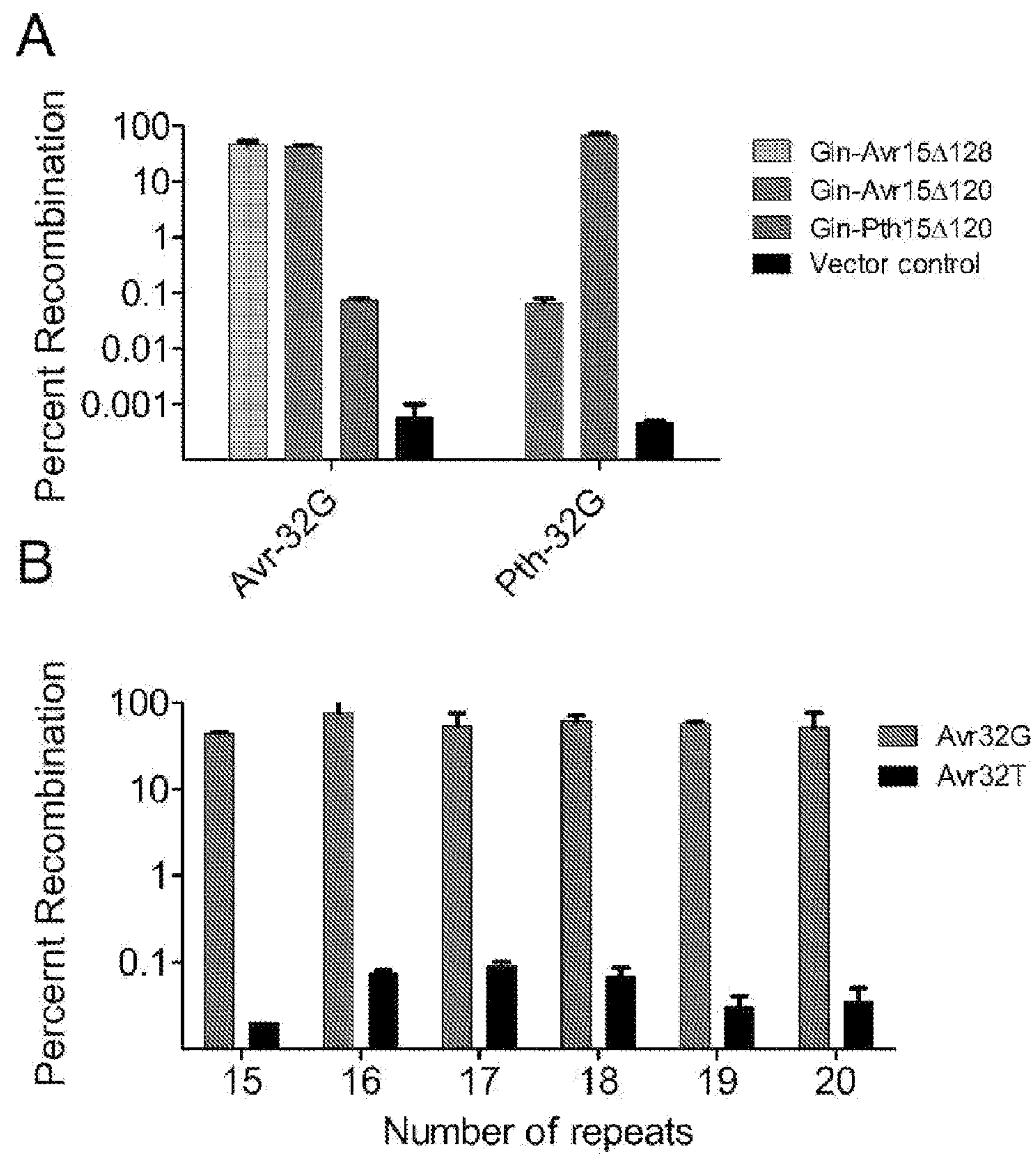


FIG. 4

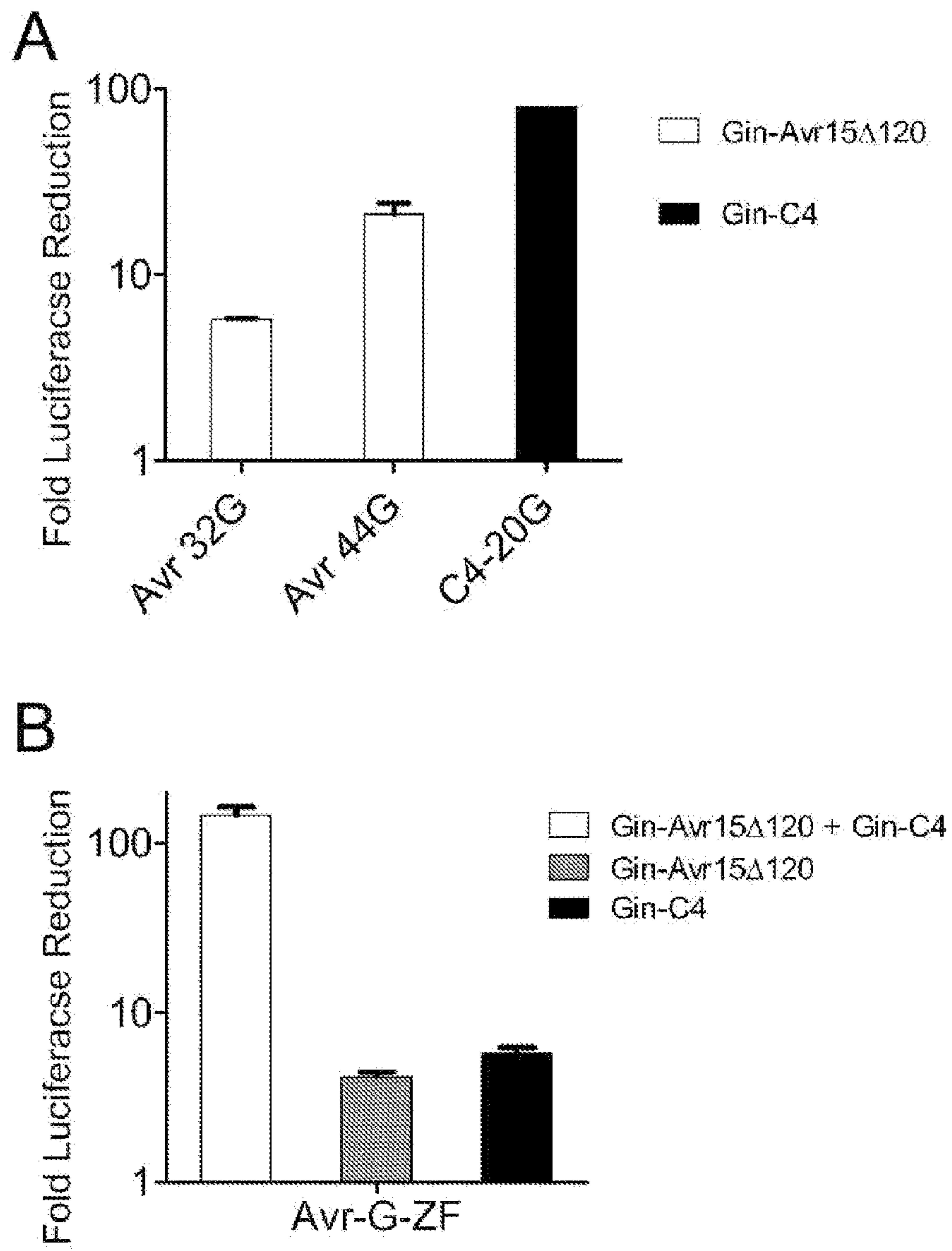


FIG. 5

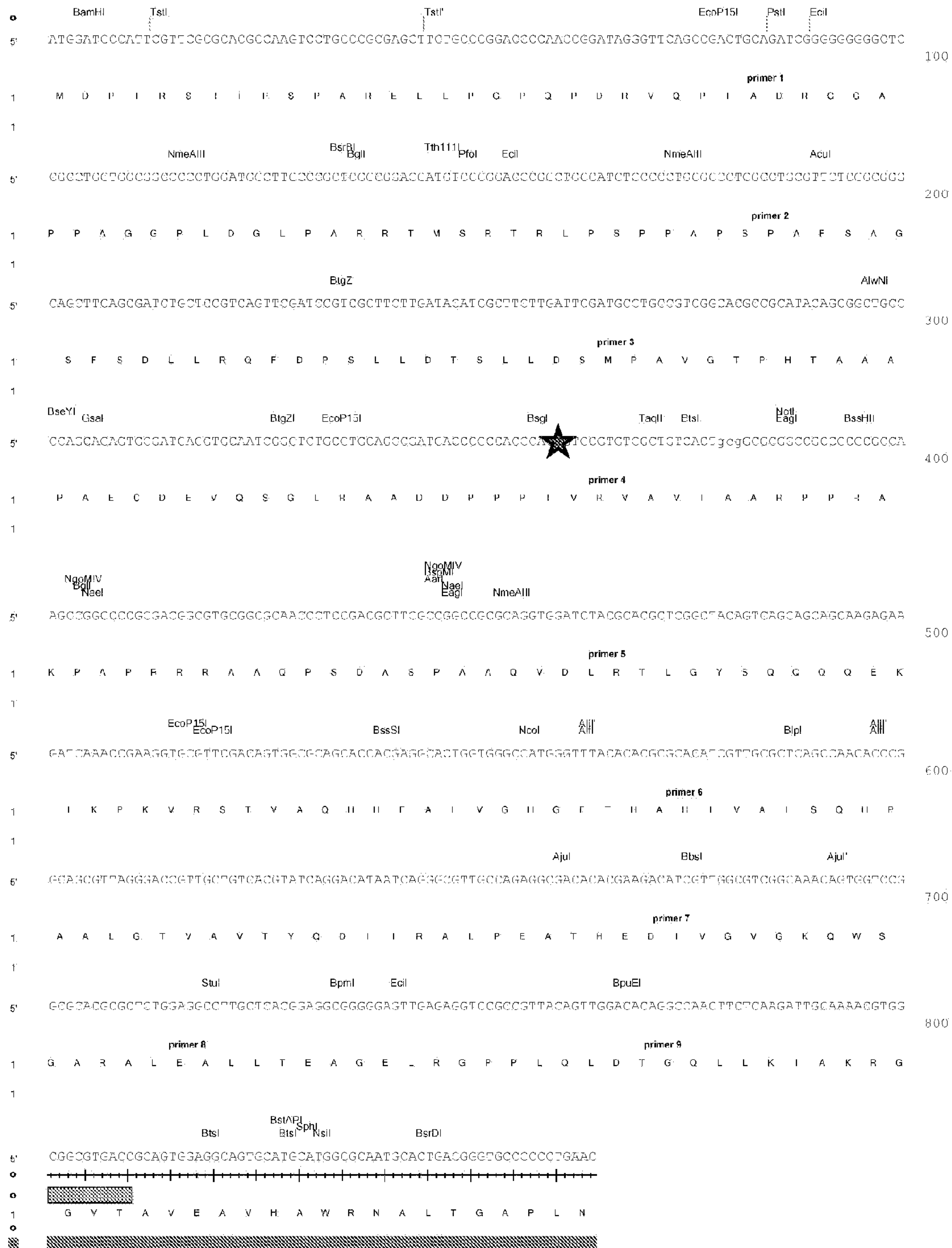


FIG. 6

AviXa7	NI HG NI NI NS HD NN HD HD HD NS N* N* HD HD NS NS NN NN NI NG NN NI N* NS N*
Target Sequence	A T A A C C C C T C C A A C C A G G T G C T A A
Synthetic	NI NG NI NI NI HD HD HD HD NG HD HD NI NI HD HD NI NN NN NG NN HD NG NI NI

FIG. 7

TALE and TALER sequences**AvrXa7 protein:**

MDPIRSRTPSPARELLPGPQPDRVQPTADRGGAPPAGGPLDGLPARRTMSRTRLPSPPAPSPA
 ESAGSFSDLLRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQSGLRADDPPPTVRVAV
 TAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGH
 GFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALLTEAGEL
 RGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVVAIASNIGGKQALET
 VQRLLPVLCQDHGLTPDQVVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNIG
 GKQALETVQRLLPVLCQAHGLTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTPAQVV
 AIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALETQRLLPVLCQDHG
 LTPDQVVVAIANNNGGKQALETQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVVAIASHDGGKQAL
 ETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASN
 GGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALETVQRLLPVQRLLPVLCQDHG
 LTQDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALETVQRLLP
 VLCQDHGLTPDQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALE
 TVQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIANN
 NGGKQALETVQRLLPVLCQDHGLTPAQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQ
 VVAIASNNGGSKQALETVQRLLPVLCQDHGLTPDQVVVAIANNNGGKQALETVQRLLPVLCQ
 DHGLTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVVAIASNNGGKQALETVQRL
 LPVLCQDHGLTPNQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVVAIASNNGGKQAL
 ESIVAQLSRPDALAALTNDHLVALACLGGRPALDAVKKGLPHAPELIRRINRRIPERTSHRV
 PDLAHVVRVLGFFQSHSHPAQAFDDAMTQFEMSRHGLVQLFRRVGVTEFEARYGTLPPASQ
 RWDRILQASGMKRAKPSPTSAQTPDQASLHAFADSLERDLDAAPSPMHEGDQTRASSRKRSR
 SDRAVTGPSTQQSFVVRVPEQQDALHLPLSWRVKRPRTTRIGGGLPDPGTPIAADLAASSTVM
 WEQDAAPFAGAADDFFPAFNEEELAWLMELLPQSGSVGGTI

FIG. 8**AvrXa7 gene:**

ATGGATCCCATTCGTTTCGCGCACGCCAAGTCCTGCCCCGCGAGCTTCTGCCCCGGACCCCA
 ACCGGATAGGGTTCAGCCGACTGCAGATCGGGGGGGGGCTCCGCCTGCTGGCGGCCCC
 CTGGATGGCTTGCCCCGCTCGGCGGACGATGTCCCGGACCCGGCTGCCATCTCCCCCTGC
 GCCCTCGCCTGCGTTCTCGGCGGGCAGCTTCAGCGATCTGCTCCGTCAGTTCGATCCGTC
 GCTTCTTGATACATCGCTTCTTGATTTCGATGCCTGCCGTCGGCACGCCGCATACAGCGGC
 TGCCCCAGCAGAGTGCGATGAGGTGCAATCGGGTCTGCGTGACGCCGATGACCCGCCAC
 CCACCGTGCGTGTCGCTGTCACT_{gcg}GCGCGGCCGCCGCGCGCCAAGCCGGCCCCGCGAC
 GGCGTGCGGCGCAACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACGCACGCTC
 GGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACAGTGGCGC
 AGCACCACGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGCGCTCAGC
 CAACACCCGGCAGCGTTAGGGACCGTTGCTGTCACGTATCAGGACATAATCAGGGCGTT
 GCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGGCGCACGC
 GCTCTGGAGGCCTTGCTCACGGAGGCGGGGGAGTTGAGAGGTCCGCCGTTACAGTTGGA
 CACAGGCCAACTTCTCAAGATTGCAAAACGTGGCGGCGTGACCGCAGTGGAGGCAGTG
 CATGCATGGCGCAATGCACTGACGGGTGCCCCCCTGAACCTGACCCCGGACCAAGTGGT
 GGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTG
 CCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGCTGGCCATCGCCAGCCA
 TGGCGGCGGCAAGCAGGCGCTGGAGACGGTGC

FIG. 9

AGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCC
 ATCGCCAGCAATATTGGCGGCAAGCAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGG
 TGCTGTGCCAGGCCCATGGCCTGACCCCGGACCAGGTGCTGGCCATCGCCAGCAATATT
 GGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACC
 ATGGCC'TGACCCCGGCCAGGT'GGT'GGCCATCGCCAGCAATAGT'GGCGGCAAGCAGGC
 GCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGG
 ACCAAGTCGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGCTGCA
 GCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGCTGGCCA
 TCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGCTGCAGCGGCTGTTGCCGGT
 GCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCCACGATG
 GCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCA
 TGGCCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCG
 CTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGG
 CCAAGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTGGAGACGGTGCAG
 CGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCAT
 CGCCAGCAATAGCGGCGGCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTG
 CTGTGCCAGGACCATGGACTGACCCCGGACCAGGTGCTGGCCATCGCCAGCAATGGCG
 GCAAGCAGGCGCTGGAGACGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
 CCTGACCCCGGACCAGGTGCTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAG
 ACGGTGCAGCGGCTGTTGCCGGTACAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
 CCTGACCCAGGACCAGGTGGTGGCCATCGCCAGCCACGATGGCGGCAAGCAGGCGCTG
 GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
 AGTGGTGGCCATCGCCAGCCACGATGGCGGCAACAGGCGCTGGAGACGGTGCAGCGG
 CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAGGTGGTGGCCATCGC
 CAGCAATAGT'GGCGGCAAGCAGGCGCT'GGAGACGGT'GCAGCGGCT'GTTGCCGGT'GCTG
 TGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCCATCGCCAGCAATAGTGGCG
 GCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
 CCTGACCCCGGACCAGGTGGTGGCCATCGCCAGCAATAACGGCGGCAAGCAGGCGCTG
 GAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCA
 GGTGCTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGG
 CTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGCGCAGGTGGTGGCCATCGC
 CAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGT
 GCCAGGACCATGGCCTGACCCTGGACCAGGTGGTGGCCATTGCCAGCAATGGCGGCAG
 CAAACAGGCGCTAGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
 CTGACCCCGGACCAAGTGGTGGCCATCGCCAACAATAACGGCGGCAAGCAGGCGCTGG
 AGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAG
 GTCGTGGCCATCGCCAGCAATATTGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGC
 TGTGCGGCTGCTGTGCCAGGACCATGGCCTGACCCTGGACCAGGTGGTGGCCATCGCC
 AGCAATGGCGGCAAGCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCC
 AGGACCATGGCCTGACCCCGAACCAGGTGGTGGCCATCGCCAGCAATAGTGGCGGCAA
 GCAGGCGCTGGAGACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGA
 CCCCAGAACAGGTGGTGGCCATCGCCAGCAATGGCGGCAAGCAGGCGCTGGAGAGCAT
 TGTGCCCAGTTATCTCGCCCTGATCCGGCGTTGGCCGCGTTGACCAACGACCACCTCGT
 CGCCTTGGCCTGCCTCGGCGGACGTCCTGCCCTGGATGCAGTGAAAAAGGGATTGCCGC
 ACGCGCCGGAATTGATCAGAAGAATCAATCGCCGCAATCCCGAACGCACGTCCCATCGC
 GTTCCCGACCTCGCGCACGTGGTTCCGCTGCTTGGTTTTTTCCAGAGCCACTCCCACCA
 GCGCAAGCATTTCGATGACGCCATGACGCAGTTCGAGATGAGCAGGCACGGCTTGGTAC
 AGCTCTTTCGAGAGTGGGCGTCACCGAATTCGAAGCCCGCTACGGAACGCTCCCCCA
 GCCTCGCAGCGTTGGGACCGTATCCTCCAGGCATCAGGG

FIG. 9 (cont.)

ATGAAAAGGGCCAAACCGTCCCCTACTTCAGCTCAAACACCGGATCAGGCGTCTTTGCA
TGCATTGCGCGATTGCTGGAGCGTGACCTTGATGCGCCAGCCCAATGCACGAGGGAG
ATCAGACGCGGGCAAGCAGCCGTAAACGGTCCCGATCGGATCGTGCTGTCACCGGCCCC
TCCACACAGCAATCTTTCGAGGTGCGCGTTCCCGAACAGCAAGATGCGCTGCATTTGCC
CCTCAGCTGGAGGGTAAAACGCCCCGCGTACCAGGATCGGGGGCGGCCTCCCGGATCCT
GGTACGCCCATCGCTGCCGACCTGGCAGCGTCCAGCACCGTGATGTGGGAACAAGATGC
GGCCCCCTTCGCAGGGGCAGCGGATGATTTCCCGGCATTCAACGAAGAGGAGCTCGCAT
GGTTGATGGAGCTATTGCCTCAGTCAGGCTCAGTCGGAGGGACGATCTGA

FIG. 9 (cont.)

Gin-AvrΔ74

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPRQFDPSLLDTSLLDSMPAVGTPHTAAAPAECDEVQS
 GLRAADDPPTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQKEKIKP
 KVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQ
 WSGARALEALLTEAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLT
 PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLC
 QDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQ
 RLLPVLCQDHGLTPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
 QALETQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETQRLLPVLCQDHGLTPDQVVAI
 ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGL
 TPAQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPV
 LCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETV
 QRLLPVQRLLPVLCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAI
 ASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGL
 TPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPV
 LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALET
 VQRLLPVLCQDHGLTLDQVVAIASNNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNN
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQV
 VAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHG
 LTPNQVVAIASNSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

FIG. 10**Gin-AvrΔ87**

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPDSMPAVGTPHTAAAPAECDEVQSGLRAADDPPTV
 RVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDLRTLGYSSQQQKEKIKPKVRSTVAQHHE
 ALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
 TEAGELRGPPQLDGTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLT
 PDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQV
 VAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
 TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETQRLLPV
 LCQDHGLTPDQVVAIANNNGGKQALETQRLLPVLCQDHGLTPDQVVAIASHDGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVQRLLPV
 LCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
 GGKQALETVQRLLPVLCQDIIGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDIIGLTPDQ
 VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
 HGLTLDQVVAIASNNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
 LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNSGGKQA
 LETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
 NSGGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

FIG. 11

Gin-AvrΔ120

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPRAKPAPRRRAAQPSDASPAAQVDL
 RTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHHPAALGTVAVTYQDIIRA
 LPEATHEDIVGVGKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIARKGGVTAVEAVHA
 WRNALTGAPLNLTDPQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHGGGK
 QALETVQRLLPVLCQDHGLTPDQVVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAI
 ASNIGGKQALETVQRLLPVLCQDHGLTPAQVVVAIASNSGGKQALETVQRLLPVLCQDHGLT
 PDQVVVAIASHDGGKQALETQRLLPVLCQDHGLTPDQVVVAIANNNGGKQALETQRLLPVL
 CQDHGLTPDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALET
 VQRLLPVLCQDHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSG
 GKQALETVQRLLPVLCQDHGLTPDQVVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNGGKQALETVQRLLPVQRLLPVLCQDHGLTQDQVVVAIASHDGGKQALETVQRLLPVLC
 CQDHGLTPDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALET
 VQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNNG
 GKQALETVQRLLPVLCQDHGLTPDQVVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQV
 VAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVVAIASNGGSKQALETVQRLLPVLCQDH
 GLTPDQVVVAIANNNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNIGGKQALETVQRLL
 PVLCQDHGLTLDQVVVAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVVAIASNSGGKQALE
 TVQRLLPVLCQDHGLTPNQVVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

FIG. 12**Gin-AvrΔ120***

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSTVRVAVTAARPPHAVAGPAAQVDLRTLGSYQQQQE
 KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHHPAALGTVAVTYQDIIRALPEATHEDIVGV
 GKQWSGARALEALLTEAGELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLN
 LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHGGGKQALETVQRLLP
 VLCQDHGLTPDQVVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVVAIASNIGGKQALE
 TVQRLLPVLCQDHGLTPAQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHD
 GGKQALETQRLLPVLCQDHGLTPDQVVVAIANNNGGKQALETQRLLPVLCQDHGLTPDQ
 VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPVLCQ
 DHGLTPAQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALETVQ
 RLLPVLCQDHGLTPDQVVVAIASNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNGGKQA
 LETVQRLLPVQRLLPVLCQDHGLTQDQVVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNSGGKQALETVQRLLPVLCQD
 HGLTPDQVVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNNGGKQALETVQRL
 LPVLCQDHGLTPDQVVVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVVAIASNIGGKQA
 LETVQRLLPVLCQDHGLTLDQVVVAIASNGGSKQALETVQRLLPVLCQDHGLTPDQVVVAIAN
 NNGGKQALETVQRLLPVLCQDHGLTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHGLTLD
 QVVVAIASNGGKQALETVQRLLPVLCQDHGLTPNQVVVAIASNSGGKQALETVQRLLPVLCQD
 HGLTPNQVVVAIASNGGKQALESIVAQLSRPDPALAALTNDHLVALACLG

FIG. 13

Gin-AvrΔ147

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPASPAAQVDLRTLGYSSQQQEQEKIKPKVRSTVAQHHE
 ALVGHGFTHAHIVALSQHPAALGTVAVTYQDIIRALPEATHEDIVGVGKQWSGARALEALL
 TEAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVAIASNIGG
 KQALETVQRLLPVLCQDHGLTPDQVVAIASHGGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGL
 TPAQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETLQRLLPV
 LCQDHGLTPDQVVAIANNNGGKQALETLQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHD
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVQRLLPV
 LCQDHGLTQDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALE
 TVQRLLPVLCQDHGLTPDQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNS
 GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQ
 VVAIANNNGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQD
 HGLTLDQVVAIASNNGGSKQALETVQRLLPVLCQDHGLTPDQVVAIANNNGGKQALETVQR
 LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTLDQVVAIASNNGGKQA
 LETVQRLLPVLCQDHGLTPNQVVAIASNSGGKQALETVQRLLPVLCQDHGLTPNQVVAIAS
 NNGKQALESIVAQLSRPDPALAAALTNDHLVALACLG

FIG. 14**GinAvr15Δ128-synthetic protein**

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGD TLVV
 WKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIERT
 MAGLAAARNKGRIGGRPRKSGSGSPALRPPRAKPAARRRAAQPSDASPAAQVDLRTLGYSSQ
 QQEQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHIITALPEATHED
 IVGVGKQWSGARALEALLTDAGELRGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNALTG
 APLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQ
 RLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQ
 ALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
 HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTP
 DQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
 QDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALET
 VQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDG
 GKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVV
 AIASNIGGKQALESIVAQLSRPDPALAAALTNDHLVALACLGPKKKRKV

FIG. 15

Gin-Avr15Δ128-synthetic DNA:

ATGCTGATTGGCTATGTAAGGGTATCAACAAATGACCAGAATACAGACCTGCAACGAA
 ACGCTCTTGTGTTGTGCAGGATGTGAACAAATATTTGAAGATAAATTAAGCGGAACAAGG
 ACAGACCGACCGGGATTAAAACGCGCTTTAAAGCGCCTTCAAAAAGGTGACACACTGG
 TTGTCTGGAAACTGGATCGCCTCGGGCGAAGCATGAAACATTTGATTTCTCTCGTAGGG
 GAATTACGAGAGCGAGGGATTAAATTTTCGCAGTCTTACTGACAGTATTGATACGTCATC
 TCCAATGGGGCGTTTTTTCTTCTACGTTATGGGTGCCCTGGCTGAAATGGAACGAGAACT
 AATTATCGAGCGAACGATGGCTGGACTTGCTGCCGCCAGAAATAAAGGCCGTATTGGAG
 GTCGCCCCGCGTAAATCGGGGTCTGGATCCCCCGCGCGGCCGCCGCGCCAAGCCGGCC
 CCGCGACGGCGTGCTGCGCAACCCTCCGACGCTTCGCCGGCCGCGCAGGTGGATCTACG
 CACGCTCGGCTACAGTCAGCAGCAGCAAGAGAAGATCAAACCGAAGGTGCGTTCGACA
 GTGGCGCAGCACCAAGAGGCACTGGTGGGCCATGGGTTTACACACGCGCACATCGTTGC
 GCTCAGCCAACACCCGGCAGCGTTAGGGACCGTCGCTGTCACGTATCAGCACATAATCA
 CGGCGTTGCCAGAGGCGACACACGAAGACATCGTTGGCGTCGGCAAACAGTGGTCCGG
 CGCACGCGCCCTGGAGGCCTTGCTCACGGATGCGGGGGAGTTGAGAGGTCCGCCGTTAC
 AGTTGGACACAGGCCAACTTGTGAAGATTGCAAAACGTGGCGGCGTGACCGCAATGGA
 GGCAGTGCATGCATCGCGCAATGCACTGACGGGTGCCCCCCTGGAGCTGACTCCGGACC
 AAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCG
 GCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCG
 CCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCT
 GTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCG
 GCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGG
 CCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG
 AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 GTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
 TGTGCGCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
 AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
 GCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGG
 CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
 CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
 AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
 TGTGCGCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCC
 AGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
 GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGG
 CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGC
 CTGACTCCGGACCAAGTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCG
 AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAA
 GTGGTGGCTATCGCCAGCCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGC
 TGTGCGCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCC
 AGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGT
 GCCAGGACCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGG
 CAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCCGGCGTTGGCCG
 CGTTGACCAACGACCACCTCGTTCGCCCTGGCCCTGCCCTCGGCCCAAGAAGAAGCGCAAG
 GTGTAG

FIG. 16

GinAvr15Δ128-synthetic protein:

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKR
ALKRLQKGDTLVWVKLDRLGRSMKHLISLVGELRERGINFRSLTDSI
DTSSPMGRFFFYVMGALAEMERELIERTMAGLAAARNKGRIGGRPR
KSGSGSPALRPPRAKPAPRRRAAOPSDASPAAQVDLRTLGYSOQQOE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVTYQHI
ITALPE
ATHEDIVGVGKQWSGARALEALLTDAGELRGPPLQ (-1)
LDTGQLVKIAKRGGVTAMEAVHASRNALTGAPLN (0)T
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG (1)A
LTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHG (2)T
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG (3)A
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG (4)A
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG (5)A
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (6)C
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (7)C
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (8)C
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (9)C
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (10)C
LTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHG (11)T
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (12)C
LTPDQVVVAIAS**HD**GGKQALETVQRLLPVLCQDHG (13)C
LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG (14)A
LTPDQVVVAIASNIGGKQALE (15)A
SIVAQLSRPDPALAAALTNDHLVALACLGPKKRKV

Key:

Gin Recombinase (italics)

TALE N-terminal framework (double underline)

RVD (DNA binding domains) (emboldened)

TALE C-terminal framework (dotted underline)

FIG. 17

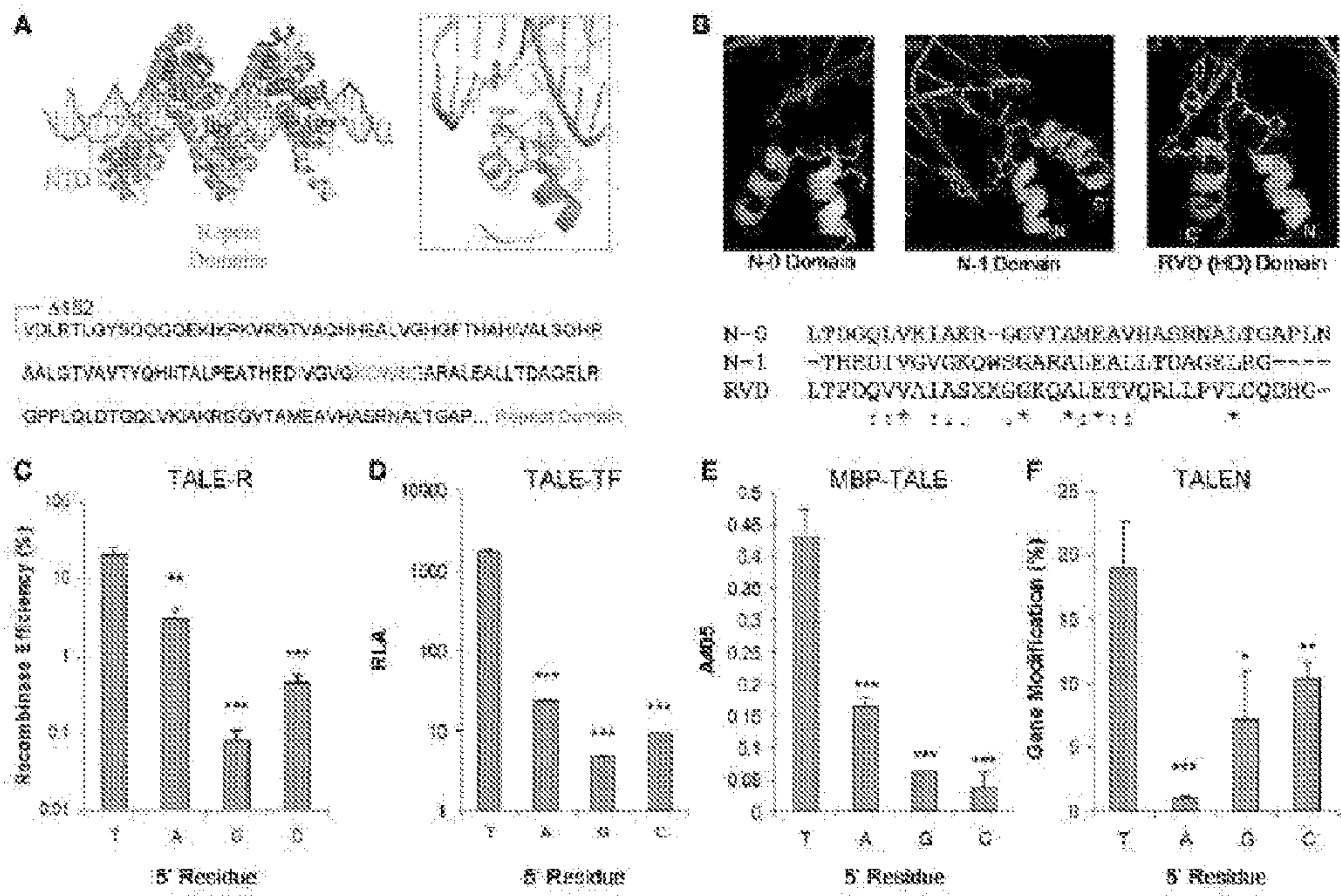


FIG. 18

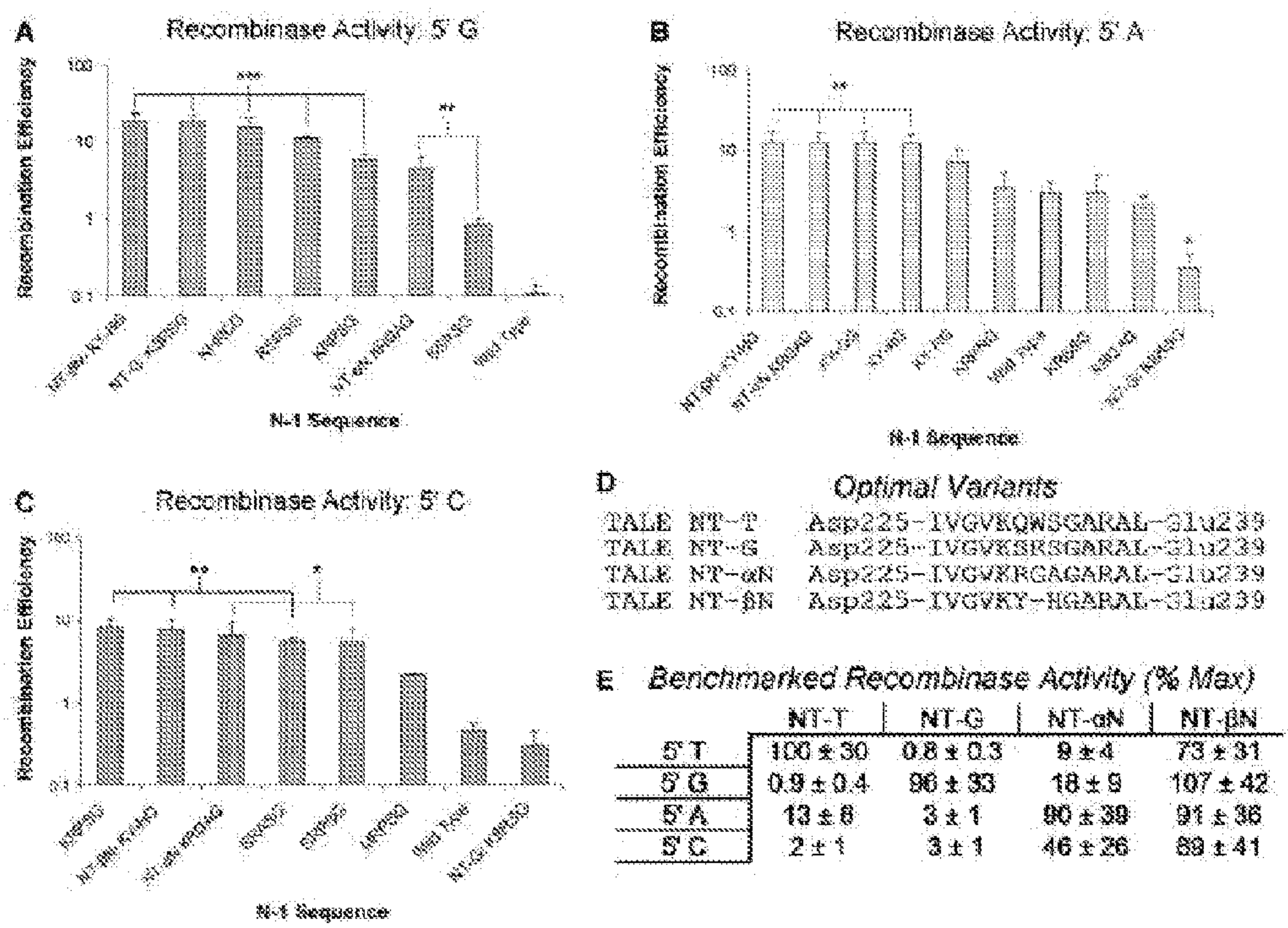


FIG. 19

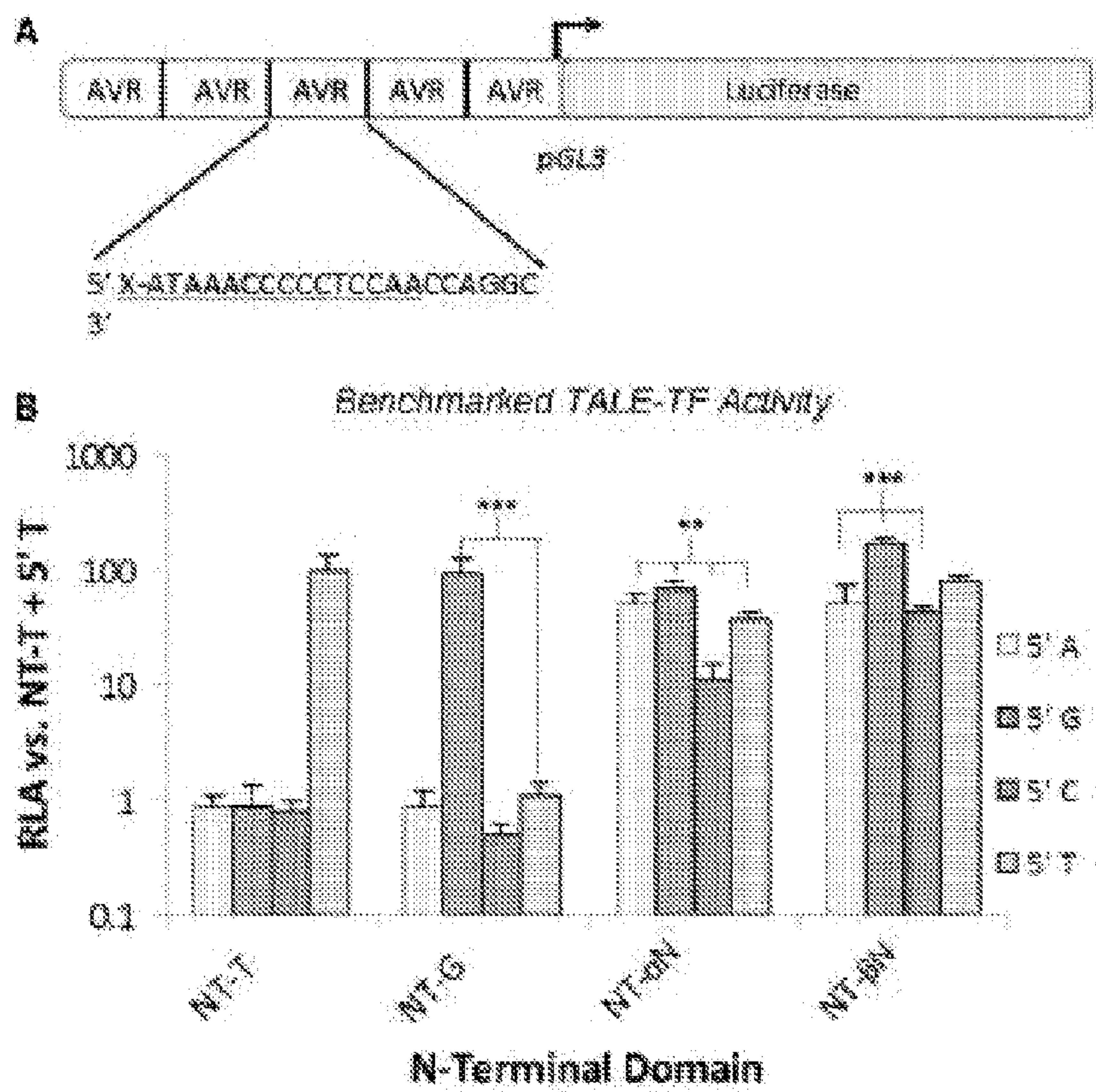


FIG. 20

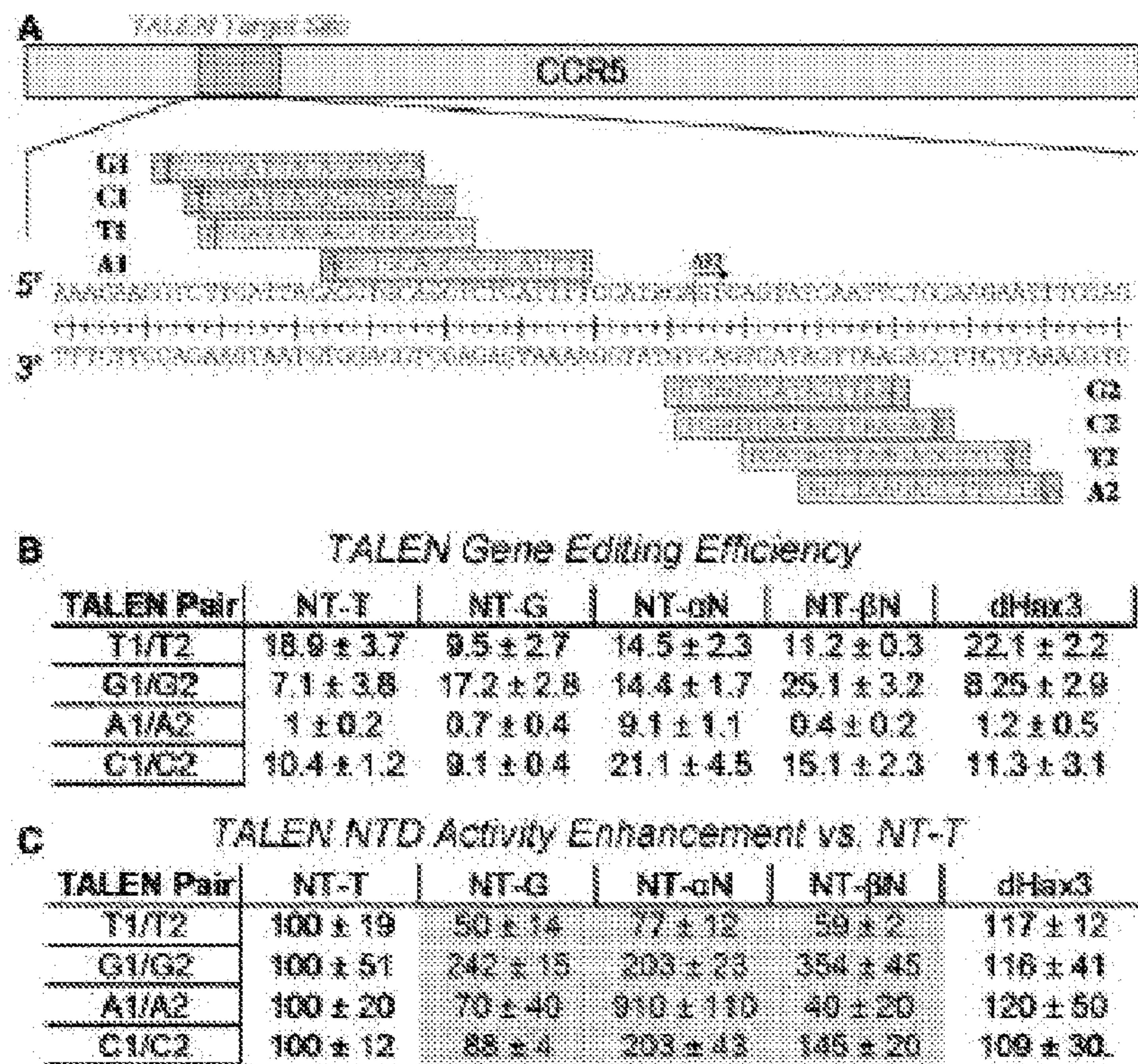


FIG. 21

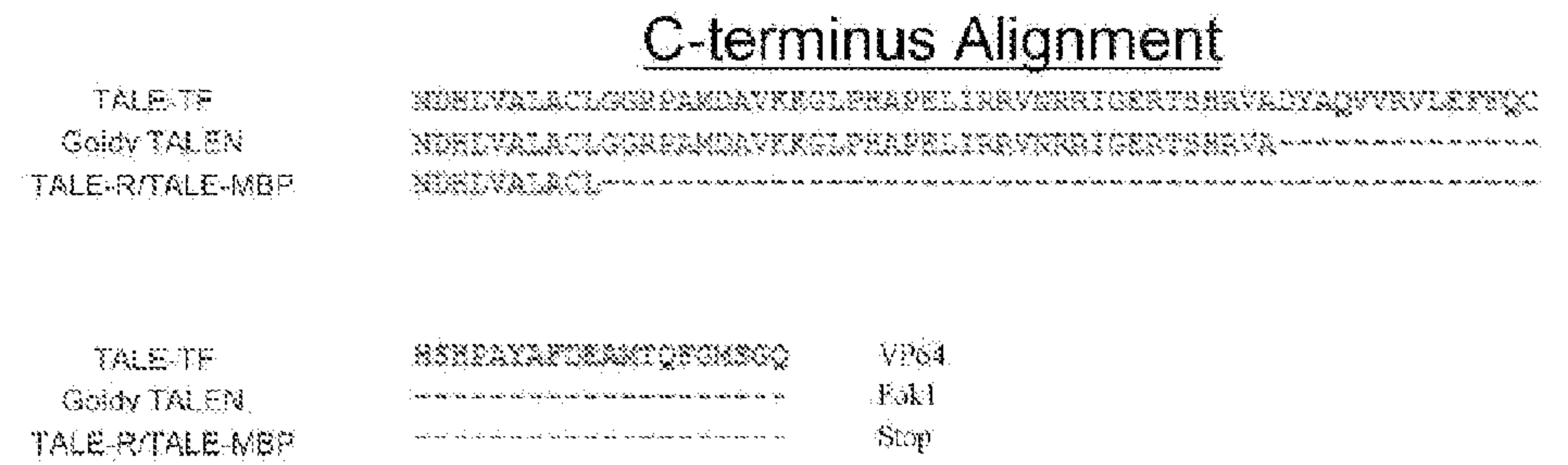
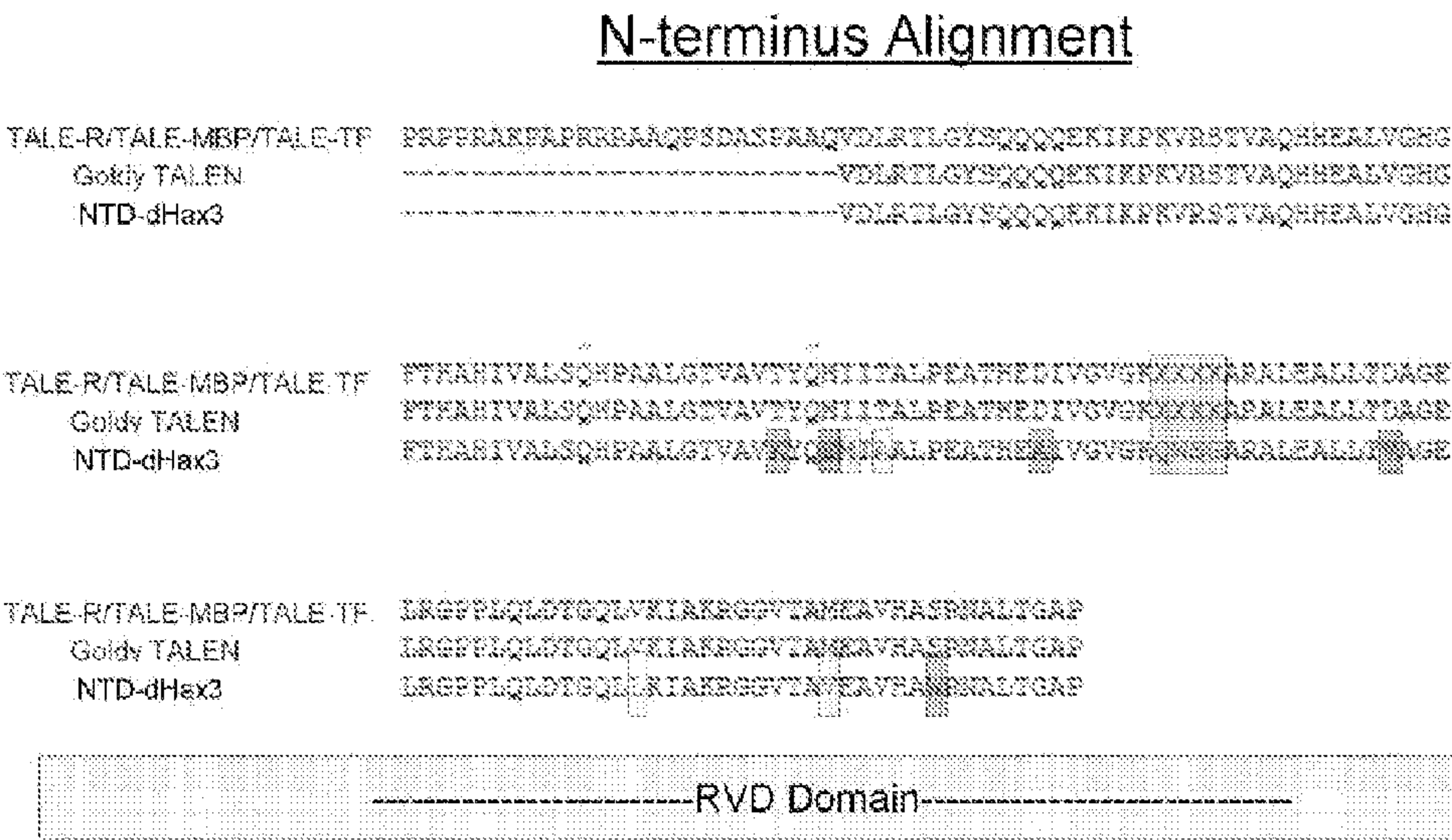


FIG. 22

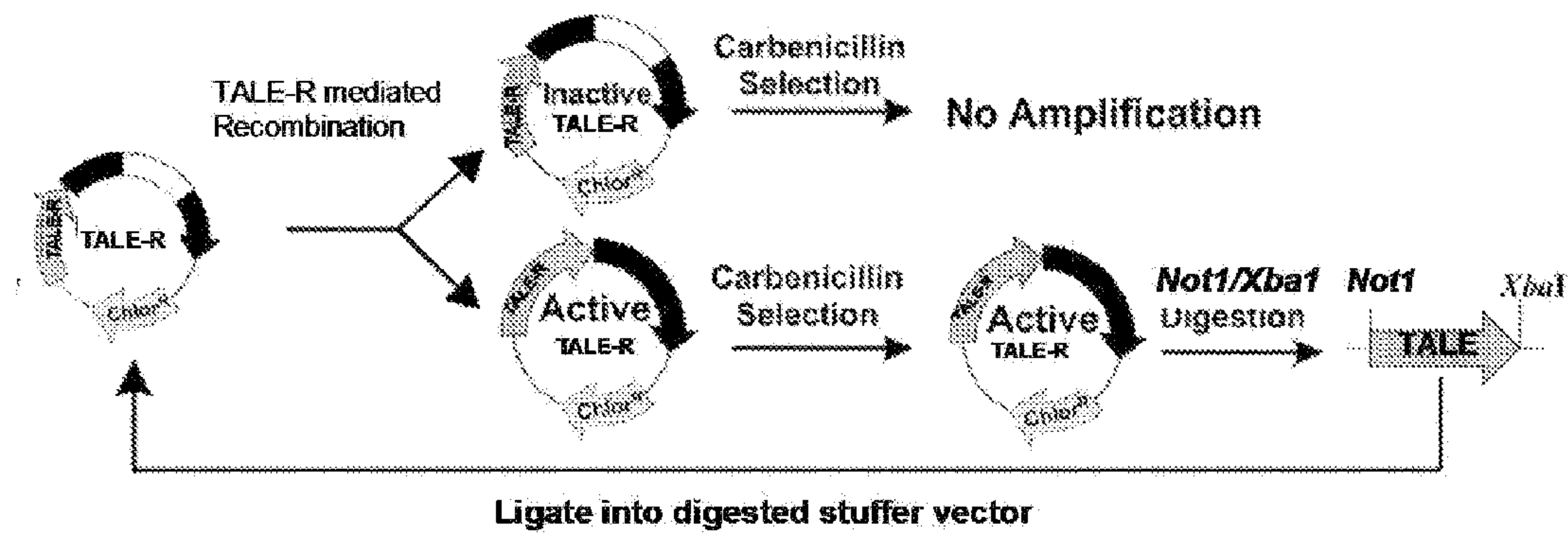


FIG. 23

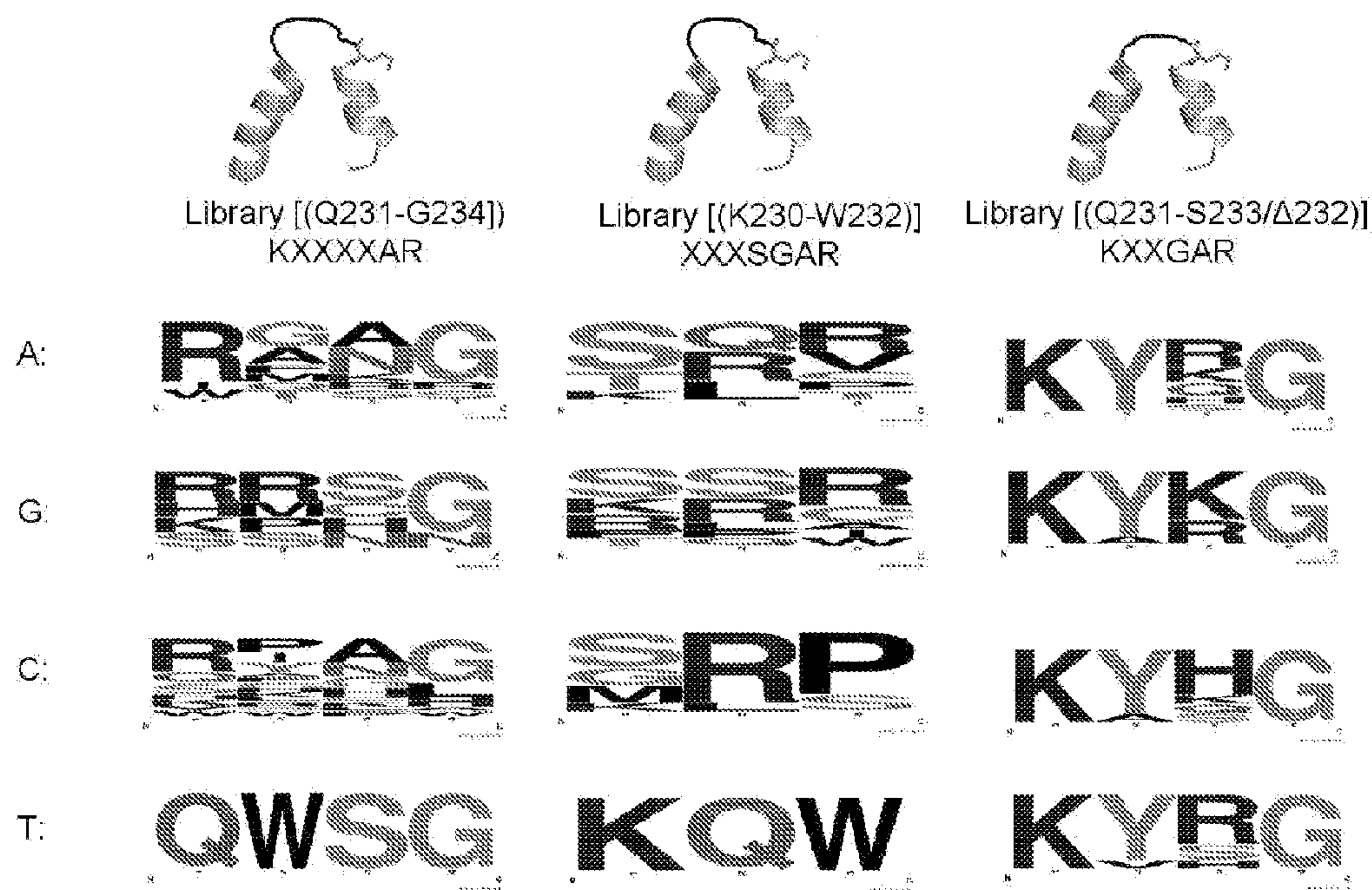


FIG. 24

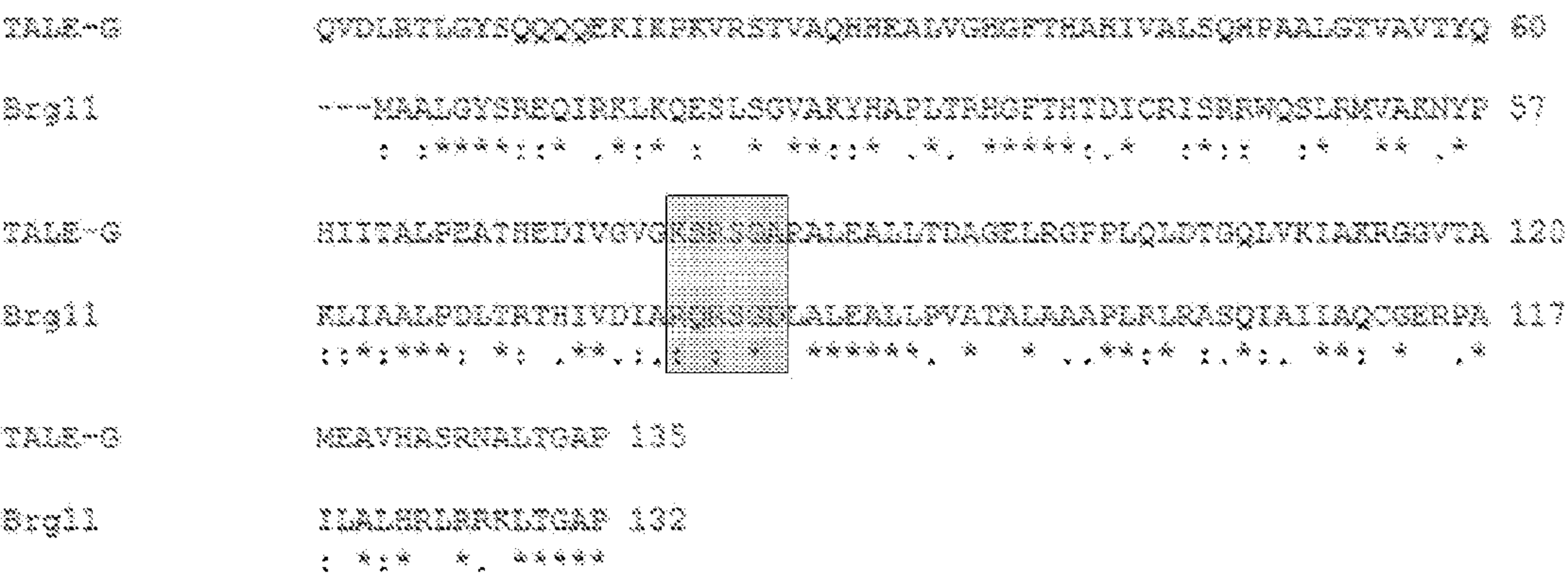


FIG. 25

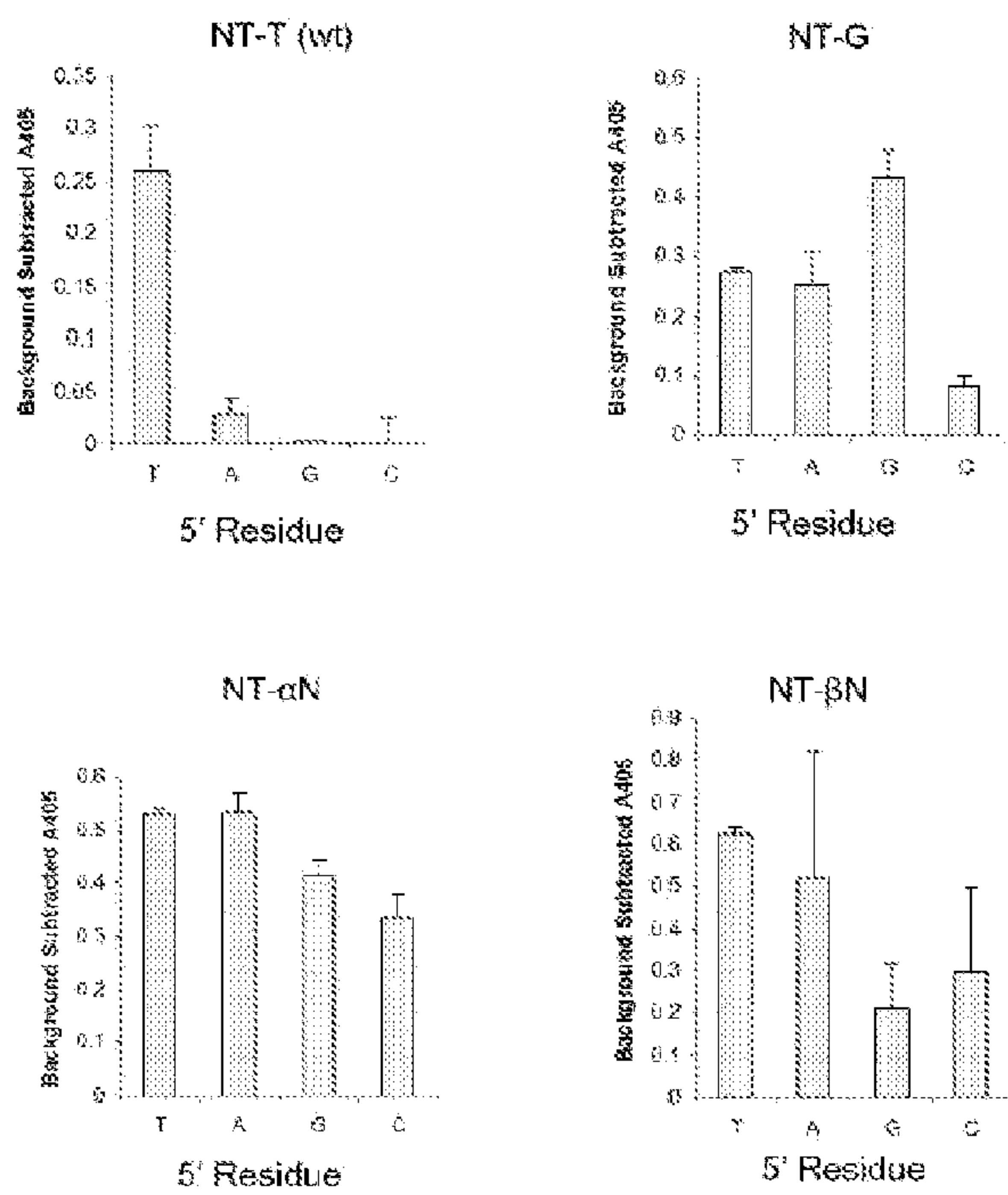
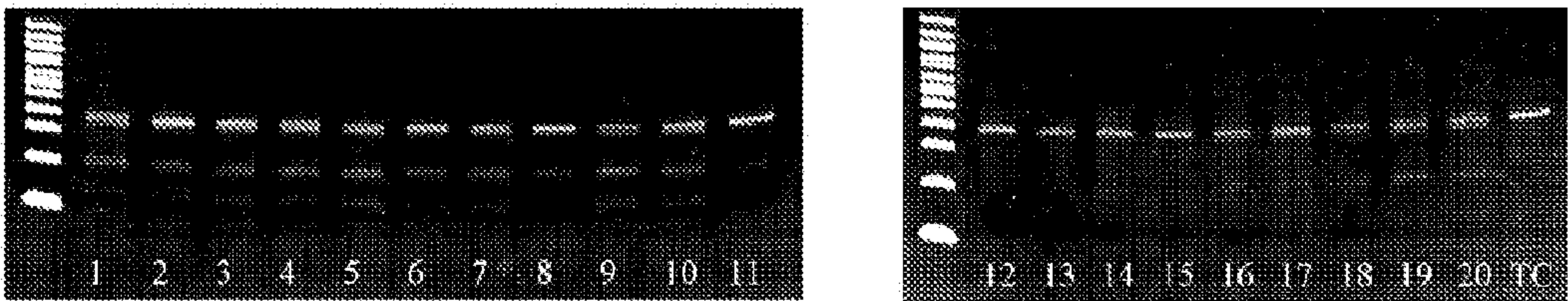


FIG. 26



TALEN Pair	NTD	% Editing	Indels	TALEN Pair	NTD	% Editing	Indels
1: T1/T2	NT-T	16.6%		12: A1/A2	NT-G	0.7%	
2: T1/T2	NT-G	7.5%		13: A1/A2	NT- βN	0.4%	
3: T1/T2	NT-βN	11.3%		14: A1/A2	NT- αN	9.1%	(3/30)
4: T1/T2	NT-αN	12.9%		15: A1/A2	NT-dHax3	1.2%	
5: T1/T2	NT-dHax3	20.1%		16: C1/C2	NT-T	8.6%	
6: G1/G2	NT-T	4.4%		17: C1/C2	NT-G	8.9%	
7: G1/G2	NT-G	15.2%	(9/30)	18: C1/C2	NT- βN	13.4%	(9/30)
8: G1/G2	NT- αN	4.9%	3x Ave = 14.4%	19: C1/C2	NT- αN	17.9%	(8/30)
9: G1/G2	NT- βN	28.9%	(10/30)	20: C1/C2	NT-dHax3	9.1%	
10: G1/G2	NT-dHax3	6.2%					
11: A1/A2	NT-T	1%					

FIG. 27



FIG. 28

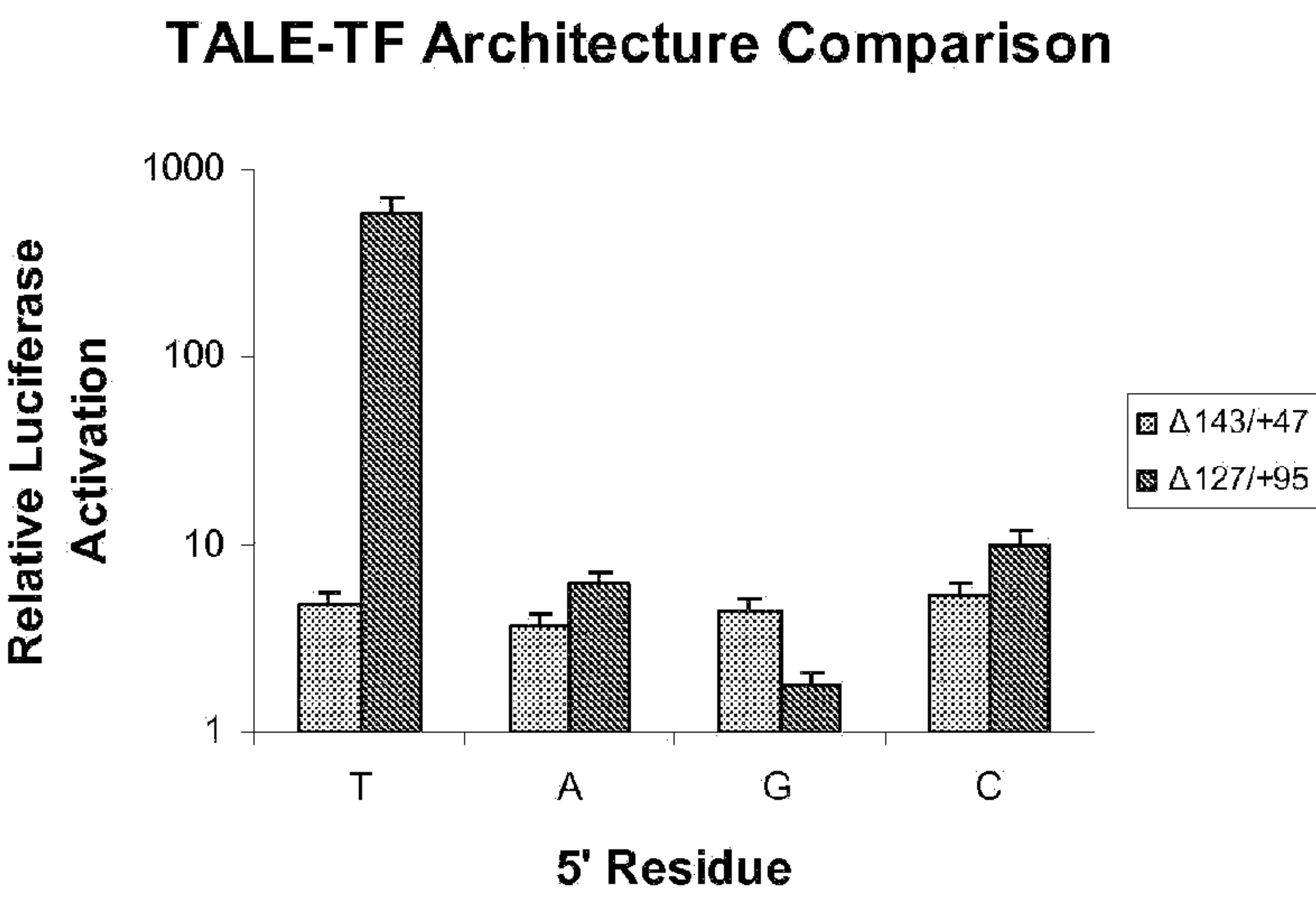


FIG. 29

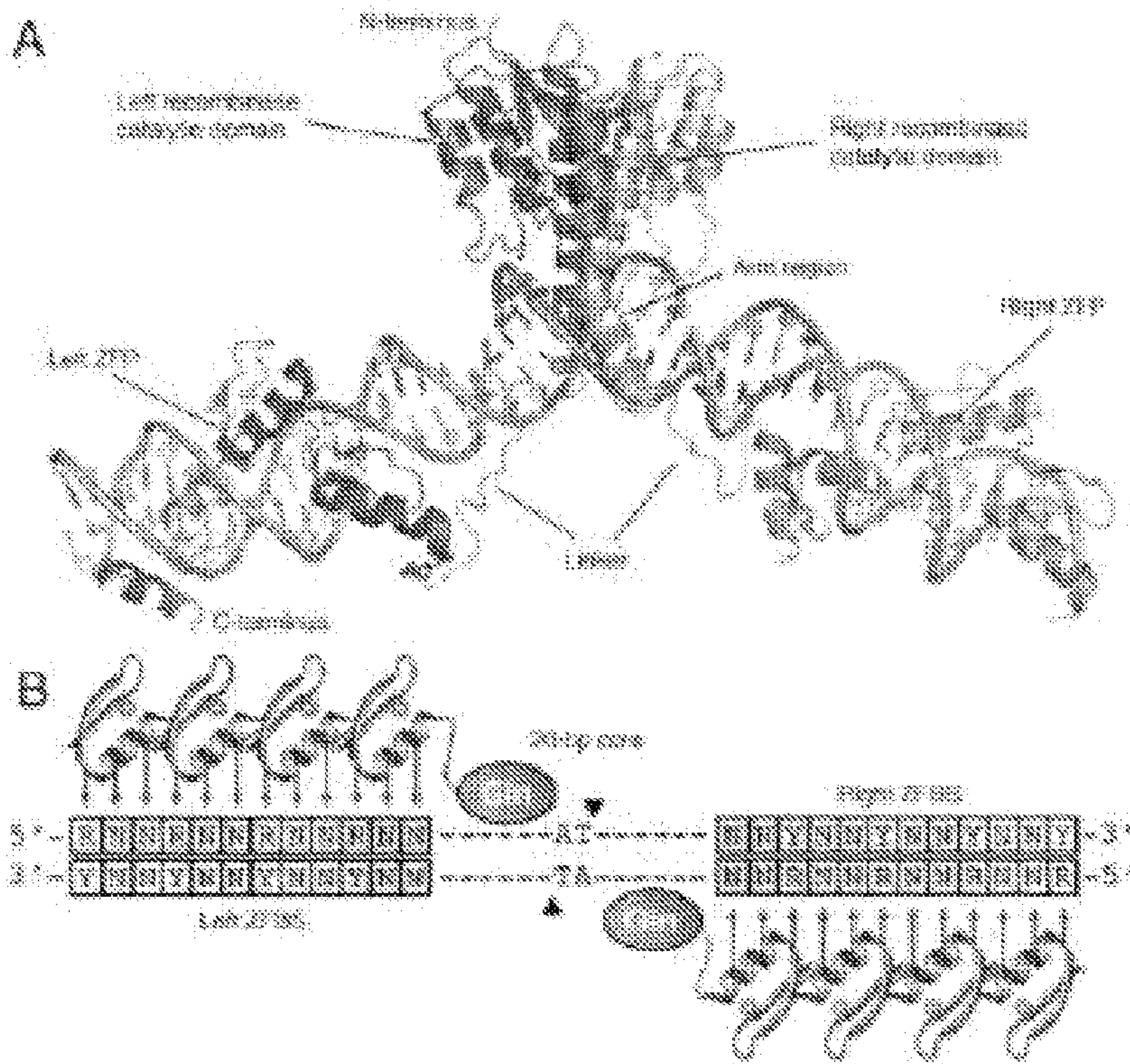


FIG. 30

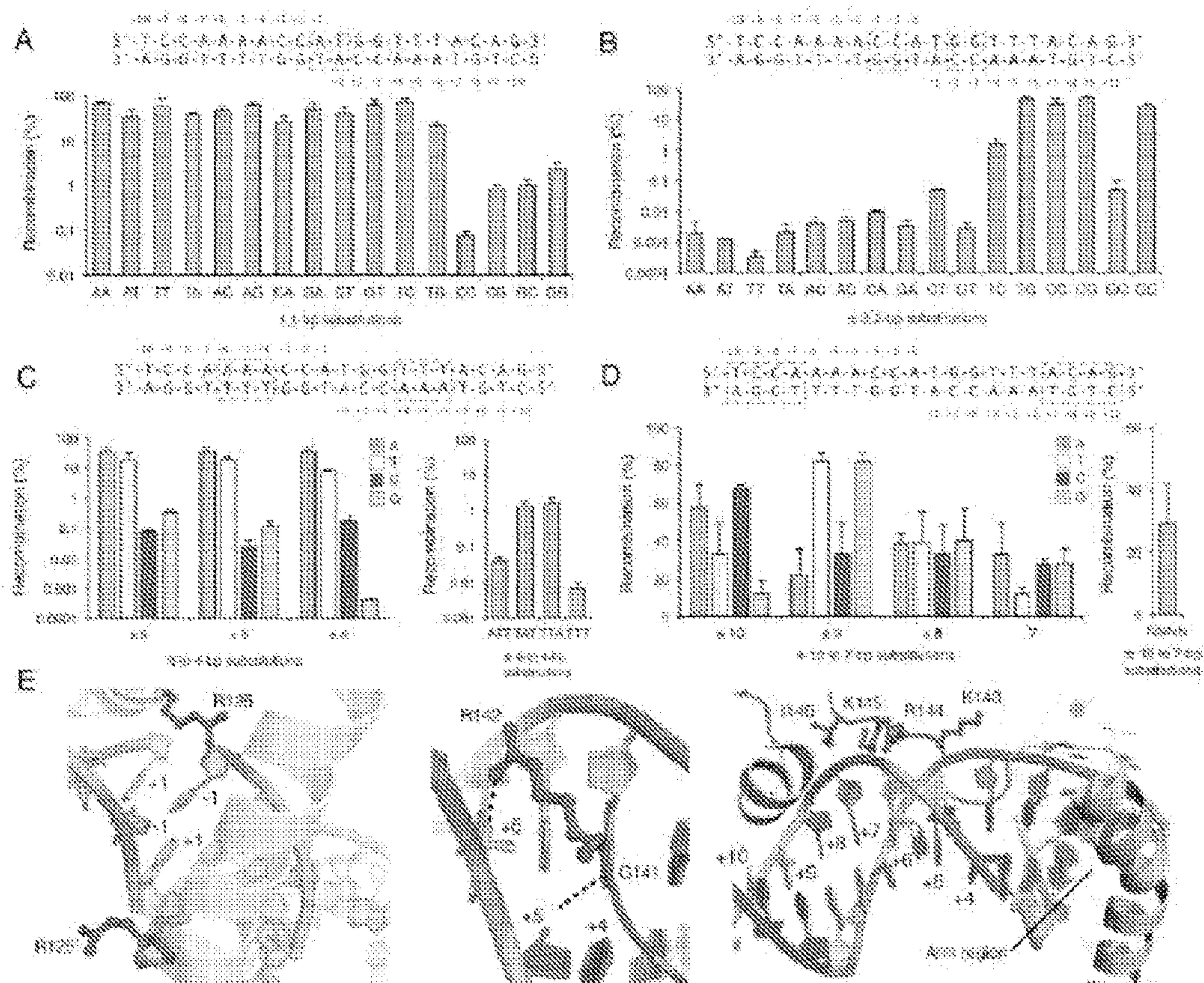


FIG. 31

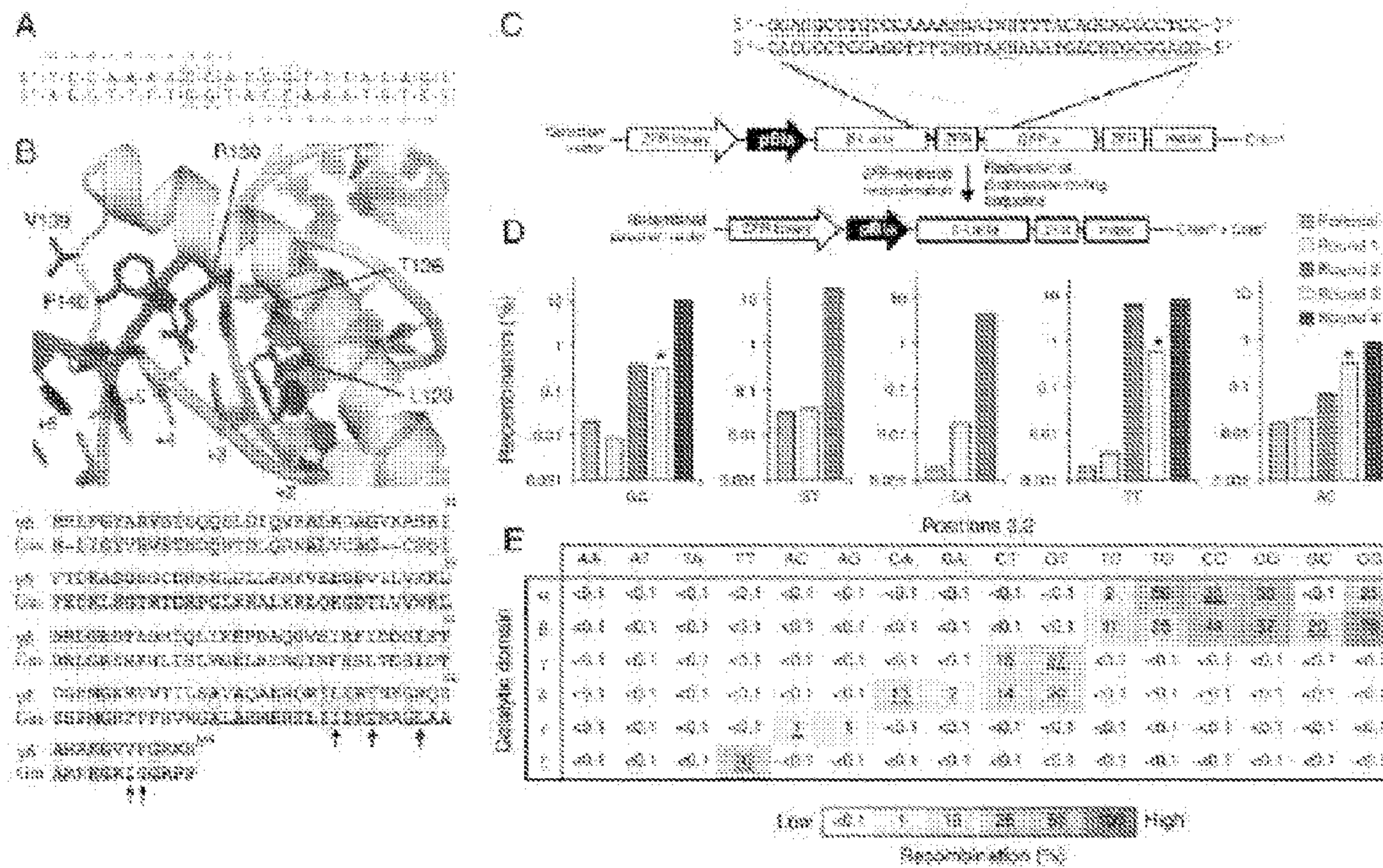


FIG. 32

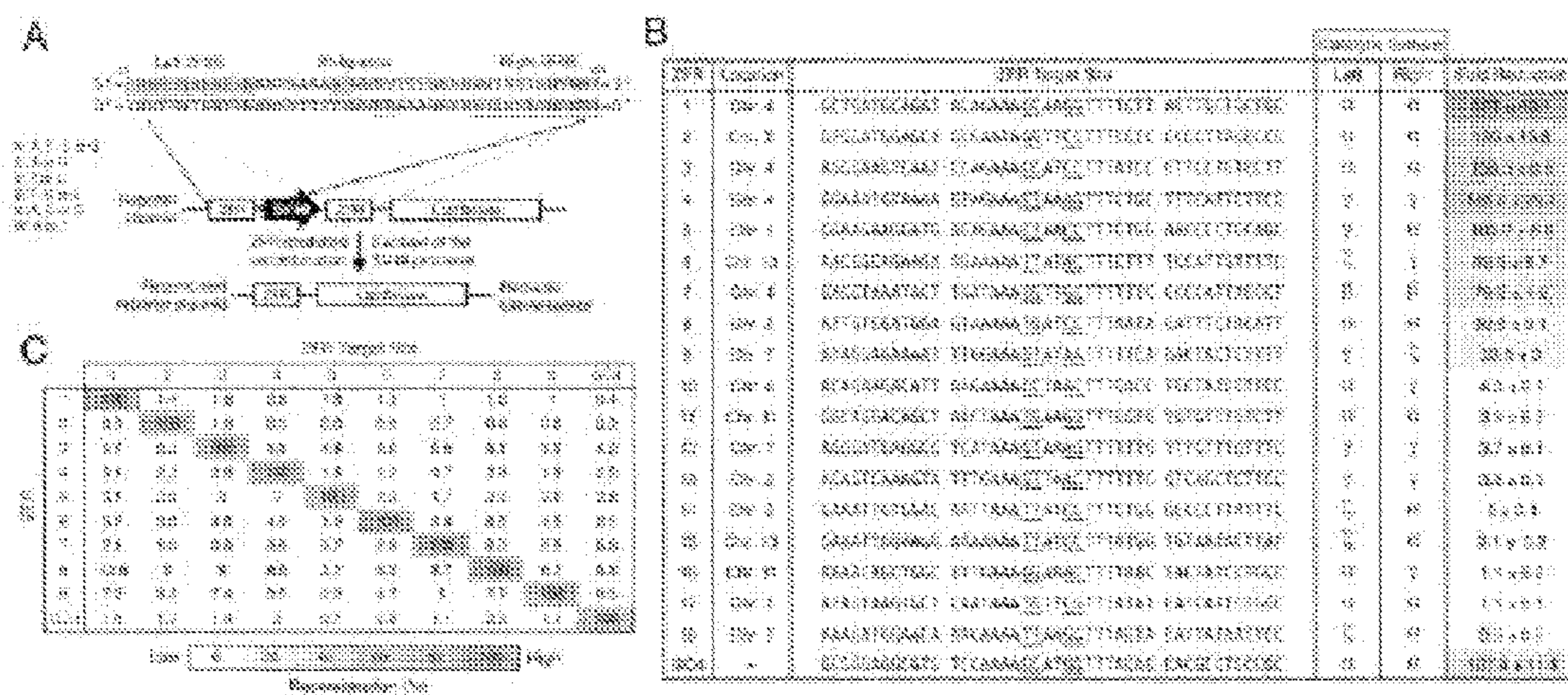


FIG. 33

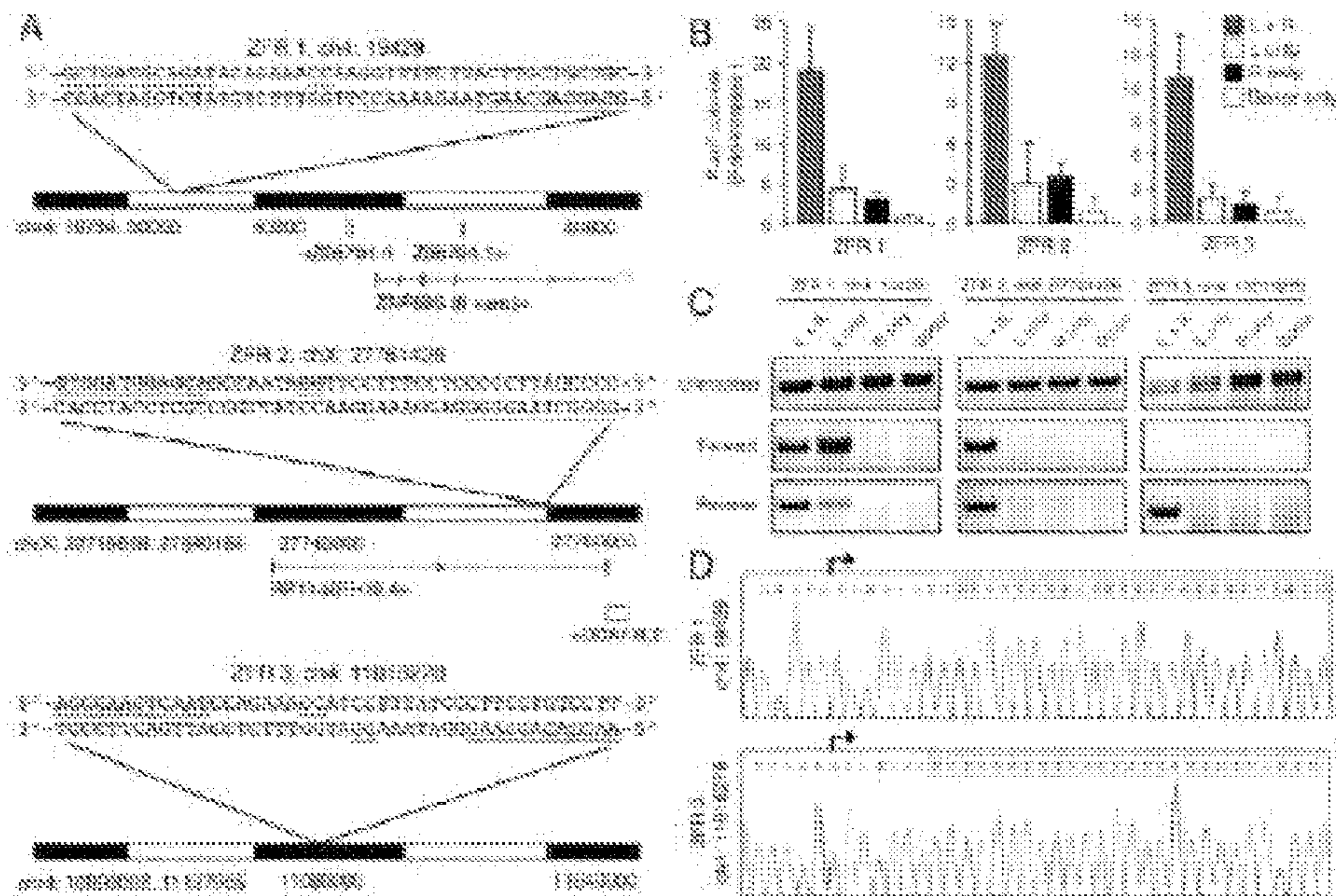


FIG. 34

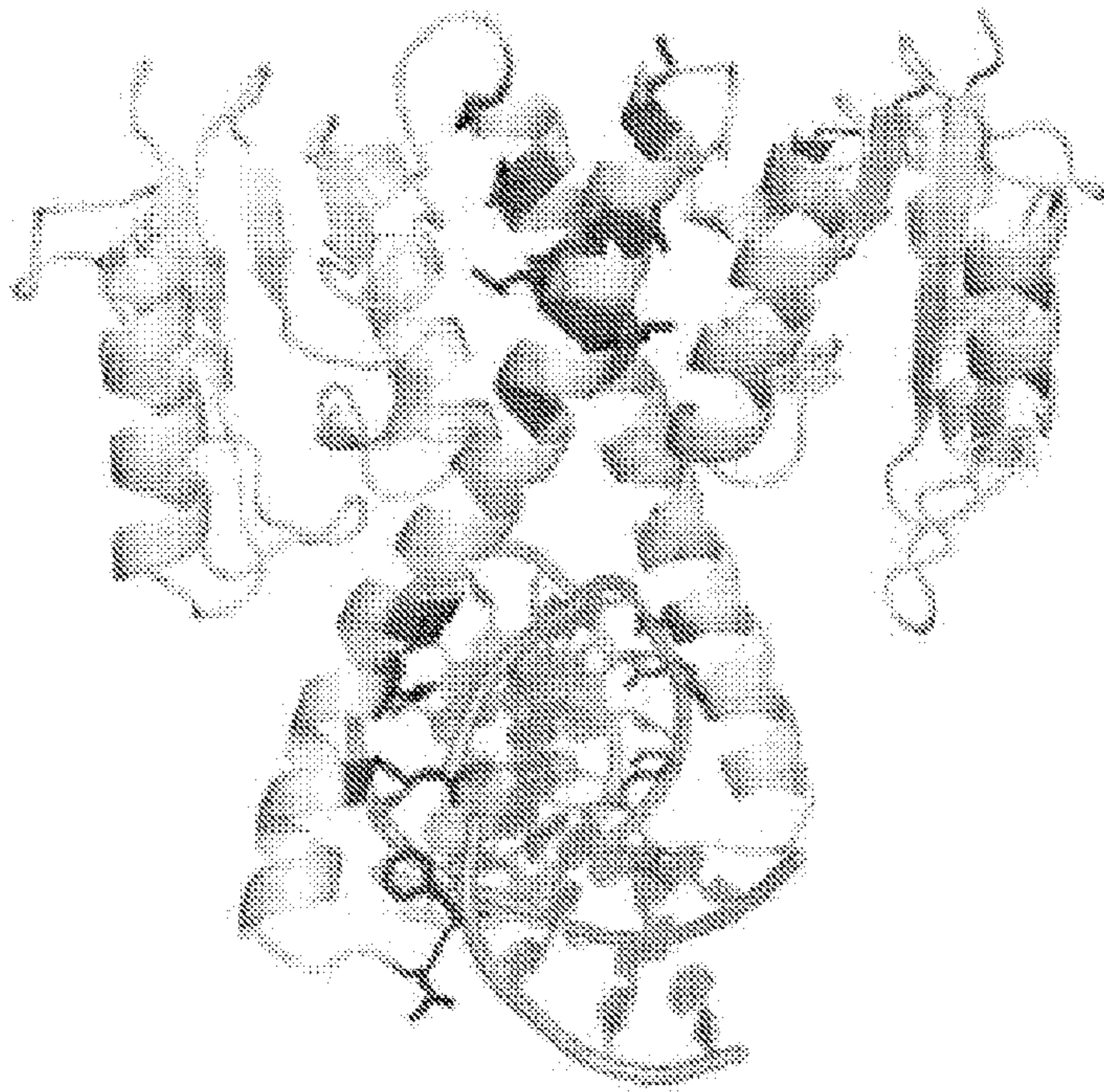


FIG. 35

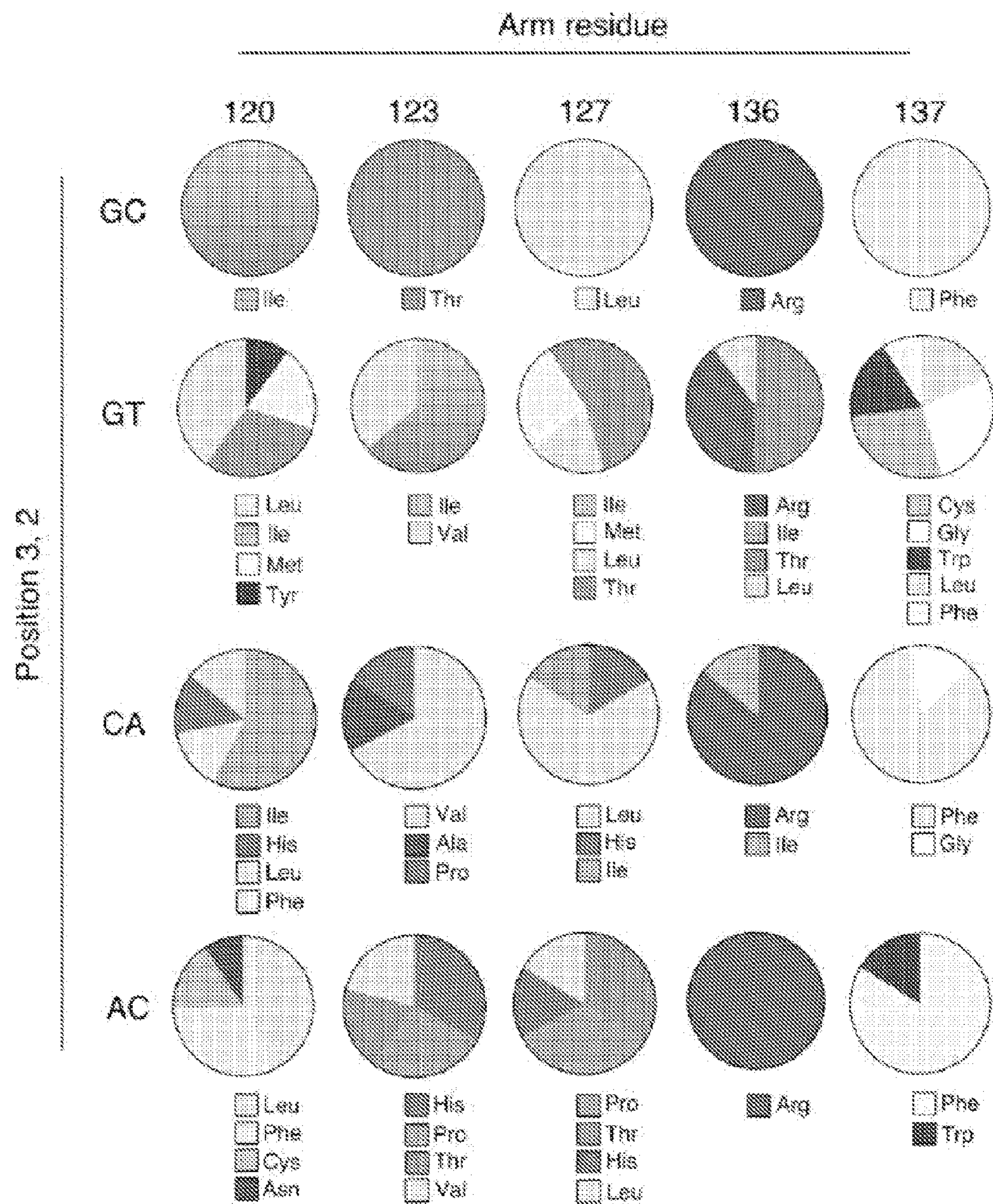


FIG. 36

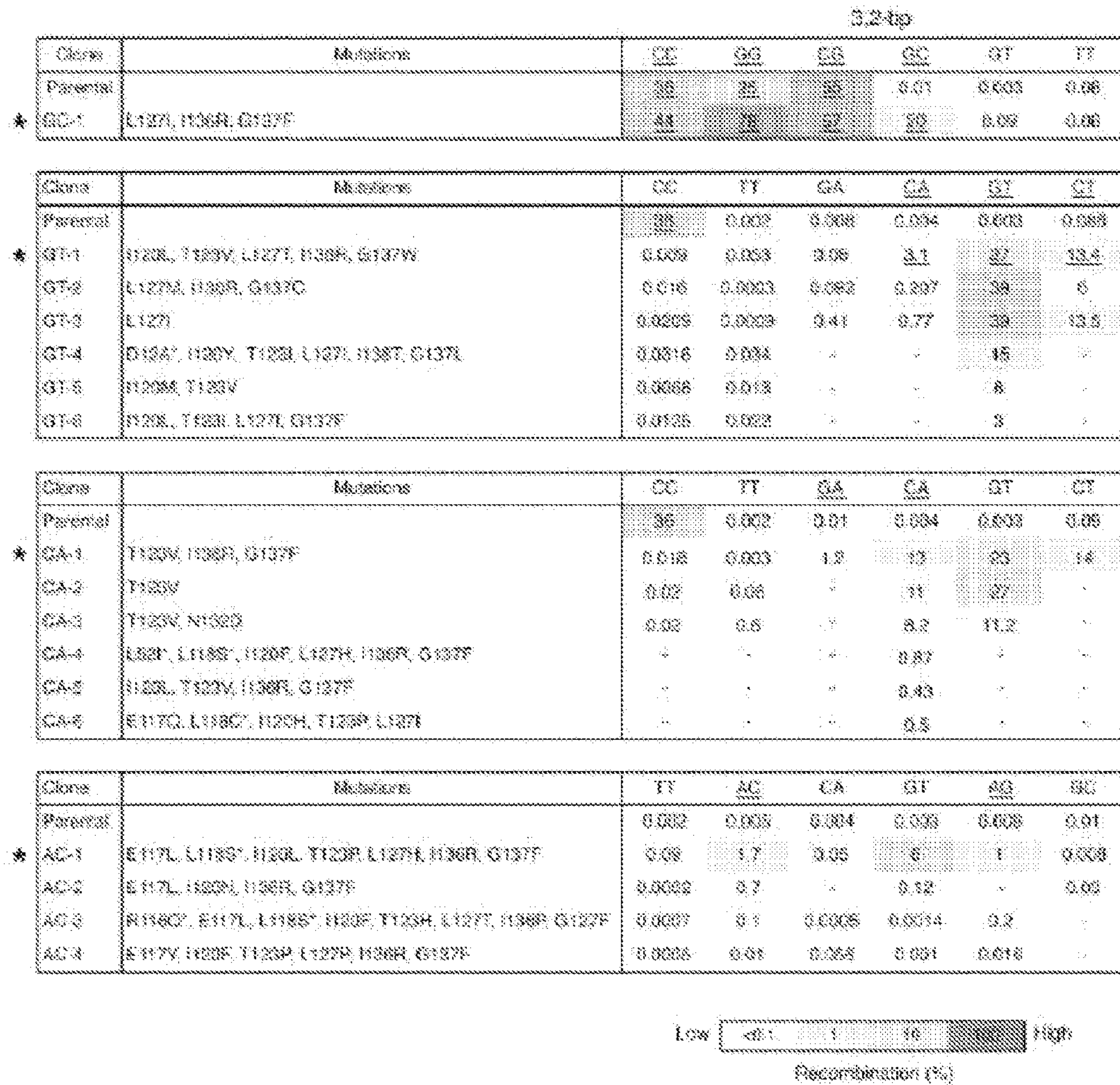


FIG. 37

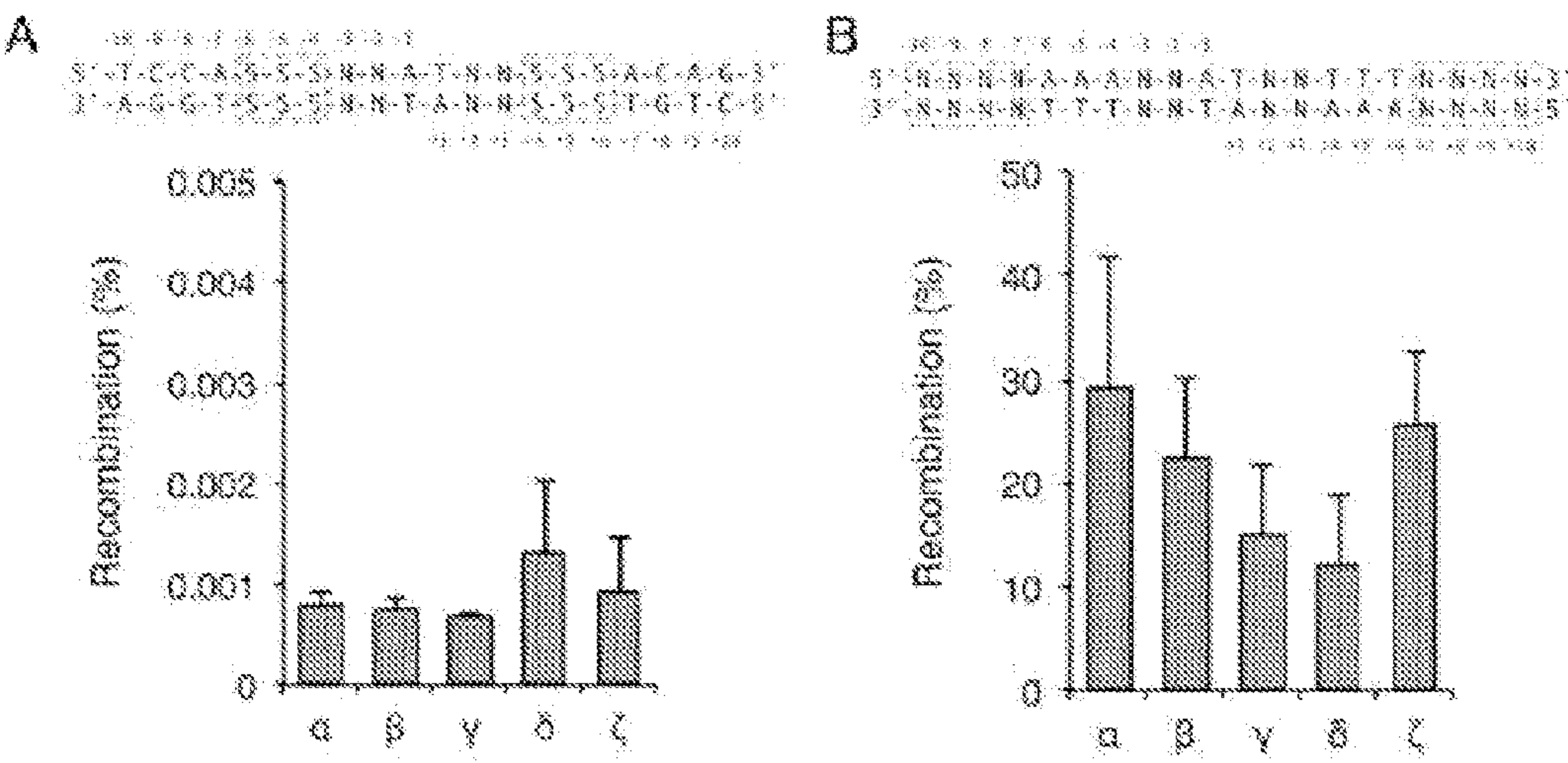


FIG. 38

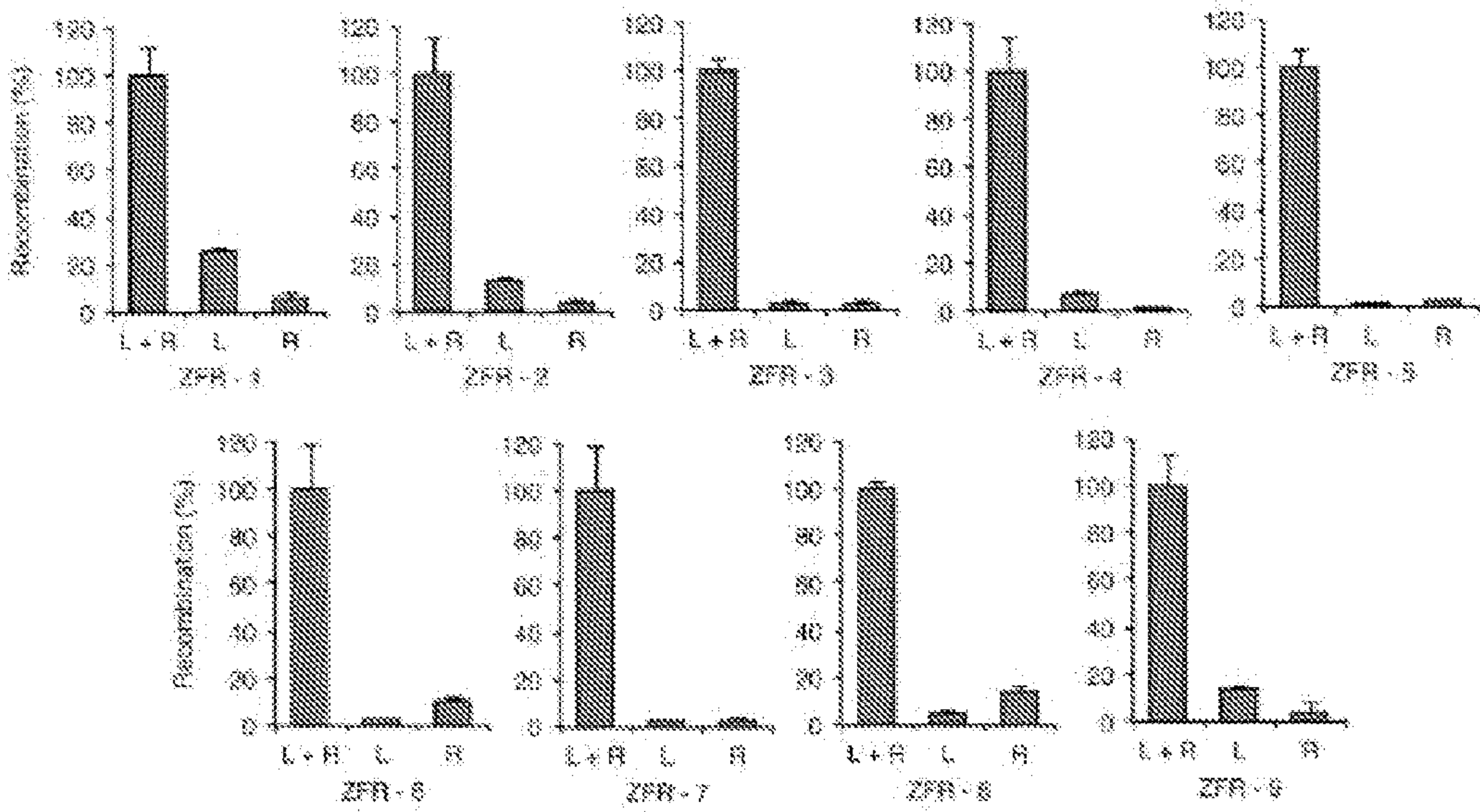


FIG. 39

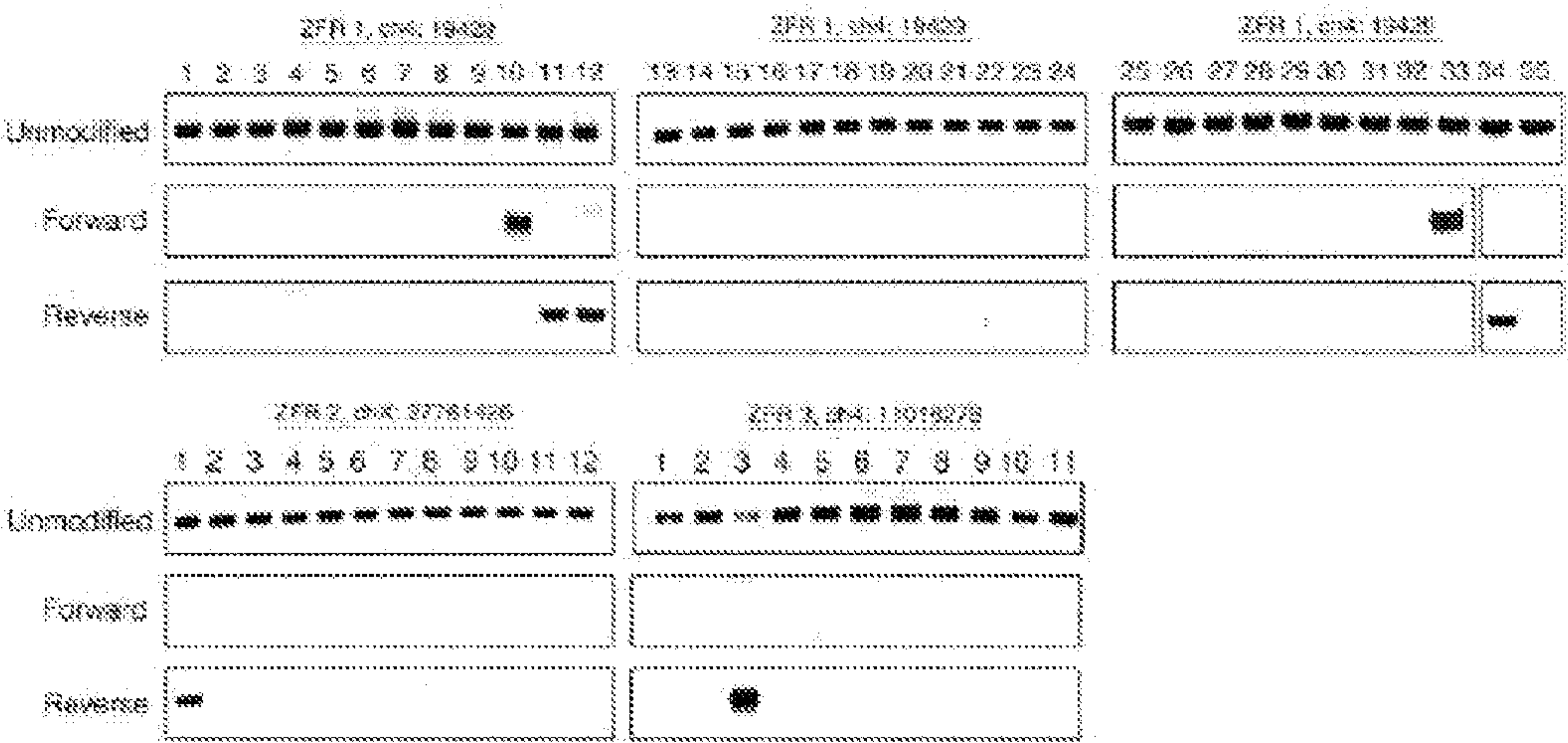


FIG. 40