(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 June 2022 (23.06.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/129438 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/62* (2006.01)
C12N 15/10 (2006.01) *C12N 15/113* (2010.01)
C12N 15/90 (2006.01)

(21) International Application Number:
PCT/EP2021/086348

(22) International Filing Date:
16 December 2021 (16.12.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20214696.5 16 December 2020 (16.12.2020) EP
21209719.0 22 November 2021 (22.11.2021) EP

(71) Applicant: **UNIVERSITAT POMPEU FABRA** [ES/ES]; Plaza de la Mercè 10-12, 08002 Barcelona (ES).

(72) Inventors: **GUELL CARGOL, Marc**; c/o UNIVERSITAT POMPEU FABRA, Plaza de la Mercè 10-12, 08002 Barcelona (ES). **SANCHEZ-MEJIAS GARCIA, Avencia**; c/o UNIVERSITAT POMPEU FABRA, Plaza de la Mercè 10-12, 08002 Barcelona (ES). **PALLARES MASMITJA, Maria**; c/o UNIVERSITAT POMPEU FABRA, Plaza de la Mercè 10-12, 08002 Barcelona (ES). **IVANCIC DJERMANOVIC, Dimitrie**; c/o UNIVERSITAT POMPEU FABRA, Plaza de la Mercè 10-12, 08002 Barcelona (ES). **RAHMEH, Amal**; c/o UNIVERSITAT POMPEU FABRA, Plaza de la Mercè 10-12, 08002 Barcelona (ES).

(74) Agent: **ICOSA**; 83 avenue Denfert-Rochereau, 75014 Paris (FR).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** PROGRAMMABLE TRANSPOSASES AND USES THEREOF

(57) **Abstract:** The present disclosure provides efficient and precise programmable gene delivery technology based on a composition comprising (i) a first protein comprising or consisting of a site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence; or a nucleic acid construct encoding said first protein; and (ii) a second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein; wherein said transposase is a modified hyperactive PiggyBac.

WO 2022/129438 A1

PROGRAMMABLE TRANSPOSASES AND USES THEREOF

TECHNICAL FIELD

The invention relates to the field of gene editing and gene therapy.

BACKGROUND ART

Many diseases such as cancer, developmental disorders, and some infections have genetic and epigenetic aberrations in common. Gene therapy is designed to introduce genetic material into cells to target and edit the genome directly in order to correct genetically dysfunctional cells and thereby cure the associated diseases.

The gene editing toolbox has considerably expanded over the last few years has a promising tool in addition to gene therapy to repair deficient genes in order to treat disorders in subjects in need thereof.

Traditionally, gene editing is based on the design of artificial endonucleases that induce a double-strand break (DSB) into the sequence of interest in the genome[1]. Cells repair the DSB through one of two major pathways: Non-Homologous End-Joining (NHEJ) or Homology Directed Repair (HDR)[2]. Recently, editing independent on DSB has been developed. Methodologies based on directly editing DNA bases with deaminases, namely base editors (BE)[3]; and in situ replacing DNA bases with aid of a reverse transcriptase (RT), namely prime editors (PE)[4], have become available.

However, pathological genetic defects can range from a few bases to large deletions. Base editors or prime editors only target a small number of bases, and HDR-based editing scales poorly with size[5]. Methodologies based on NHEJ have been developed such as Homology Independent Targeted Integration (HITI)[6]. This methodology has been demonstrated for insertions of several kilobases but remains inefficient for very large edits[5]. While HITI might work to deliver exons, it may not be efficient enough to robustly deliver cDNAs of genes such as Dystrophin (~14kb) or Laminin-α2 (~9kb).

High precision CRISPR programmable transposons have been described in bacteria[7,8] but are not available for mammalian cells. Previous attempts of fusing zinc fingers or *Streptococcus pyogenes* Cas9 (SpCas9) to the mammalian compatible piggyBac (PB) or sleeping beauty transposases delivered systems with relatively low levels of precision[9–11]. PB system is an attractive tool for gene therapy as efficiency scales well with size[12], it is a mutation independent technology, and it works in any tissue as dependence on DNA repair mechanisms is low.

Therefore, there is still a need to develop novel systems for targeted gene delivery in mammalian cells, either in vitro, ex vivo or ex vivo.

SUMMARY OF THE DISCLOSURE

Certain programmable transposases and their use in targeted gene editing have been disclosed in WO2020250181, which content is incorporated here by reference.

The present disclosure now provides further efficient and precise programmable gene delivery technology based on a composition comprising (i) a first protein comprising or consisting of a site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence; or a nucleic acid construct encoding said first protein; and (ii) a second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein; wherein said transposase is a modified hyperactive PiggyBac. Such technology has the capability to deliver small but also large nucleic acid fragments. The inventors have tested the technology in mammalian cells and *in vivo* mouse liver and surprisingly achieved high efficiency (5-10%) of site directed integration in all of them.

In one embodiment, the composition comprises (i) a first protein comprising or consisting of a site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence; or a nucleic acid construct encoding said first protein; and (ii) a second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein; wherein said transposase is a modified hyperactive PiggyBac,

comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9.

In one embodiment, the first protein and the second protein are fused together to form a fusion protein, optionally through a linker. In one embodiment, the first protein is fused to the C terminal end of the second protein, optionally through a linker.

In one embodiment, said transposase is a modified hyperactive PiggyBac, comprising one or more amino acid mutations to increase excision activity as compared to unmodified hyperactive PiggyBac, and/or one or more amino acid mutations to decrease DNA binding activity as compared to unmodified hyperactive PiggyBac.

In one embodiment, said one or more amino acid mutations do not consist of R372A, K375A, and D450N. In one embodiment, said one or more amino acid mutations are selected among the amino acid substitutions which increase excision activity at position of M194, D450, T560, S564 S573, S592 or F594, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions M194V and/or D450N. In one embodiment, said one or more amino acid mutations are selected among the amino acid substitutions which increase excision activity at position of M194 or D450, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions M194V and/or D450N. In one embodiment, said one or more amino acid mutations are selected among the amino acid substitutions which decrease DNA binding activity at position R275, R277, R347, R372, K375, R376, E377, and/or E380, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions R275A, R277, R347S, R372A, K375A, R376A, E377A, and/or E380A. In one embodiment, said one or more amino acid mutations are selected among the amino acid substitutions which decrease DNA binding activity at position R372, K375, R376, E377, and/or E380, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions R372A, K375A, R376A, E377A, and/or E380A.

In one embodiment, the modified hyperactive PiggyBac includes the double mutations N347S and D450N, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9. In one embodiment, the modified hyperactive PiggyBac mutation comprises one of the following amino acid substitution or combination of amino acid substitutions: R372A/K375A/R376A/D450N, K375A/R376A/E377A/E380A/D450N, R372A/K375A/R376A/E377A/E380A/D450N, M194V, R376A, E377A, E380A, M194V/R372A/K375A, S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, T43I/R372A/K375A/A411T/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G; the position number corresponding to the amino acid number of the hyperactive PiggyBac of SEQ ID NO: 9, typically said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 2-8, 10-18 and 135-149.

In one embodiment, the composition further comprises a third protein comprising or consisting of a second transposase; or a nucleic acid construct encoding said third protein; wherein said second transposase is either an hyperactive PiggyBac with SEQ ID NO: 9, or a modified hyperactive PiggyBac with comprising one or more amino acid mutations as compared to the hyperactive PiggyBac with SEQ ID NO: 9. In one embodiment, the first, second and third proteins are fused together to form a triple fusion protein, optionally through a linker.

In one embodiment, the first protein comprises or consists of an RNA-guided nuclease or nickase, or a zinc finger nuclease. In one embodiment, said first protein is a nuclease protein comprising an active DNA cleavage domain and a guide RNA binding domain and having at least 80%, 90%, 95%, 99% or at least 100% identity to a Streptococcus pyogenes Cas9 (SpCas9) of SEQ ID NO: 31, Staphylococcus aureus Cas9 (SaCas9) of SEQ ID NO: 72, Cpf1 of SEQ ID NO: 74, Campylobacter jejuni Cas9 (CjCas9) of SEQ ID NO: 29, Streptococcus pyogenes Cas9 nickase (nCas9) of SEQ ID NO: 70, CasX of SEQ ID NO: 75, or Staphylococcus aureus Cas9 nickase of SEQ ID NO: 76; preferably wherein said first protein is a Cas9 protein selected from the group consisting of a Staphylococcus aureus Cas9 (SaCas9) of SEQ ID NO: 72 and Streptococcus pyogenes Cas9 (SpCas9) of SEQ ID NO: 31.

In one embodiment, the composition further comprises a guide RNA, and an exogenous nucleic acid for insertion in a genome.

In one embodiment, the transposase is fused to an RNA binding protein capable of binding to at least one specific RNA sequence comprised in the guide RNA; optionally wherein said RNA binding protein is an MS2 bacteriophage coat protein (MCP) and wherein the guide RNA comprises a MS2 RNA tetraloop binding sequence, preferably sharing at least 75% identity with SEQ ID NO: 153.

In one embodiment, the exogenous nucleic acid is a large DNA fragment, typically having a size between 5 kb and 25 kb, and more preferably between 8 kb and 20 kb.

In one embodiment, the composition is comprised in a nanoparticle.

The present invention also relates to a nucleic acid encoding any one of the fusion proteins disclosed herein, typically in the form of a messenger RNA (mRNA).

The present invention also relates to an in vitro method for site specific integration of an exogenous nucleic acid sequence into the genome of a cell, the method comprising delivering to the cell the composition of the invention, a guide RNA, and the exogenous nucleic acid.

The present invention also relates to the composition of the invention, a guide RNA, and an exogenous nucleic acid, for use in the treatment of a disease, by site-specific integration of the exogenous nucleic acid sequence into the genome of a cell.

LEGENDS OF THE FIGURES

Figure 1. Programmable transposase technology: cas9 (in red) is combined with an engineered PB transposase domain (in pink). Table of the mutants used in the experimentation with their corresponding position in PB's core model (position 563 is on the C-t which is not included in the model).

Figure 2. A Programmable transposase dependence on variants of cas9. Nuclease cas9 and PB fusion shows better results in targeted and overall insertion as opposed to dead cas9 (dcas9) or nickase cas9 (ncas9) fusions. Blue indicated targeted insertion and yellow off-target insertion. B, Programmable transposase dependence of variants of PB. Excision enhanced mutants with reduced DNA binding present the best on-target:off-target ratio (orange). On-target insertions were performed at AAVs site (green), and TRAC site (blue). C. Testing of different linkers. Linkers length and topology does not affect significantly on-target activity of Spcas9 and PB fusions.

Figure 3. Hershey reporter cell line: HEK293T cell line was engineered to contain a C-terminal fragment of a GFP preceded by one splicing acceptor and gRNAs target sites. A PB transposon was generated combining CAG promoter, N-terminal fragment of GFP followed by an splicing donor. In grey triangles, PB ITRs; SA: splicing acceptor; SD: splicing donor; Target: targeted insertion site; * insertion process disrupted ITR.

Figure 4. A Programmable transposase dependence of variants of PB. Excision enhanced mutant 450 in the context of different mutations to reduce DNA binding present the best on-target. Simultaneous mutation of R372 and R376 to A is not well tolerated. Although E377 is not involved in DNA binding, the mutation to A may be beneficial to avoid a negative charge build-up in that region upon mutation of K375 and R376 to A. B R372A/K375A decrease the integration activity of PB as a result of a decrease in binding

to target DNA (as observed for D450N as well). Testing of Off-target integration in progress.

Figure 5. Double stranded breaks and programmable DNA binding domain effects in targeted insertion. Co-localization of double stranded break and PB in the insertion site is required for efficient on-targeted insertion.

Figure 6. Sanger sequencing validation of multiple insertions (see in Fig. 2a a more comprehensive distribution measured by NGS). ITRs TTAA's are lost in the process of targeted insertion. NGG Pam is highlighted in red.

Figure 7. Insertion activity PB K375A_R376A_E377A_E380A_D450N without cas9. In order to further investigate the targeted insertion mechanism, hyPB K375A_R376A_E377A_E380A_D450N was cloned without cas9 and its insertion efficiency was tested in comparison with hyPB WT using an RFP transposon in hek293T cells. Results show no insertion activity of this mutant without fused cas9.

Figure 8. A Characterization of targeted insertion site using Guide-seq. Programmable Transposase generates irreversible insertion by inactivating ITR site by multiple indels. B Characterization of overall insertion site using Guide-seq. Only on-target insertions were detected on the TCR loci (upper panel). Sanger sequencing is shown for 4 clones (bottom panel).

Figure 9. Programmable transposase characterization of insertional profiling by Guide-seq shows that hyPB mutants in combination with Cas9 performed precise transposon insertion.

Figure 10. Benchmarking of Cas9-hyPB R372A-K375A-D450N to other targeted insertion platforms such as Cas9 induced HDR (300bp homology arms were used).

Figure 11. *in vivo* deployment of Cas9-hyPB R372A-K375A-D450N in mice liver. Relative copy number measured by qPCR is reported.

Figure 12. Programmable transposase can be engineered with different Cas variants, such as CasX, CjCas9 Cpf1 or SaCas9, some of them achieved similar results in terms of

programmable insertion at the target site as with SpCas9. Each of the Cas variant tested were targeted to the specific target region of the split GFP reporter cell line with 3 independent gRNAs.

Figure 13. Double stranded breaks, by Cas9 and a single gRNA (gRNA-TCR1 or AAVS1-3) or by nickase Cas9 and two gRNAs targeting at nearby positions (gRNA-TCR1 and AAVS1-3), and programmable DNA binding domain (ZnF) in fusion to modified hyPB (mutants R372A-K375A-D405N) results in targeted insertion. Co-localization of double stranded break and PB in the insertion site is required for efficient on-targeted insertion. This can be achieved by nuclease Cas9 or double cut by nickase Cas9.

Figure 14. Programmable transposase can be engineered as a dimer polypeptide of two hyPB domains and a Cas9 nuclease, resulting in better programmable insertion compared to Cas9-hyPB. Split GFP reporter cell line was used for the programmable insertion of split GFP transposon to the target site. The mutant of hyPB R372A-K375A-D450N has been used for the monomer or dimer fusion to Cas9. Conditions: 1-Negative control with only hyPB as insertion machinery; 2: Positive control of Cas9-hyPB R372A-K375A-D450N in pcDNA expression vector; 3: Positive control of Cas9-hyPB R372A-K375A-D450N in Lentivirus expression vector; 4: Cas9 nuclease fused to two units of hyPB R372A-K375A-D450N in C-terminal; 5: Cas9 nuclease fused to two units of hyPB R372A-K375A-D450N one in C-terminal and the other one in N-terminal.

Figure 15. Several cycles of selection of cells where programmable transposition took place allowed for the selection of best mutant combinations from a library. We identified several mutants with better enrichment, and programmable insertion capacity than Cas9-hyPB R372A-K375A-D450N when fused to Cas9.

Figure 16. On-target efficiency increases over cycles of selection. Bulk variants selected from each cycle were co-transfected with gRNA targeting AAVS1 and ½ GFP transposon into the reporter cell line. Quantity of plasmid was corrected by PB copy number to normalize for cloning efficiency.

Figure 17. (**A**) On-target efficiencies of the top selected candidates. Six individual candidates were selected based on the highest on-target activity among 96 random clones selected from the last cycle. The individual on-target activities were compared to Cas9-hyPB R372A-K375A-D450N. (**B**) Logo showing the predominant PB residues in top on-target activity variants.

Figure 18. Benchmarking of Cas9-hyPB R372A-K375A-D450N (FiCAT) to Homology-independent targeted insertion (HITI).

Figure 19. Programmable insertion activity of FiCAT R372A-K375A-D450N using four different nuclease proteins. SpCas9 is used as control for programmable insertion with gRNA-TRAC-1 only (left). Each nuclease was used with three independent gRNAs (1–3) for targeted insertion in ½ GFP reporter cell line.

Figure 20. Liver integration of minicircle luciferase transposon. Minicircle luciferase transposon, sgRNA targeting Rosa26 locus and FiCAT (Cas9-hyPB R372A-K375A-D450N) mRNA were delivered by hydrodynamic injection and luciferase signal was monitored.

Figure 21. (**a**) Editing activity by CasX (left) and Cpf1 (middle). (**b**) Editing activity by SaCas9 (left), CjCas9 (middle). Mean %of reads with indels +/- SD is shown for two technical repeats, representative image of N=3 biological replicates. SpCas9 targeting the TRAC-1 site was used for reference (right).

Figure 22. Increase of on-target efficiency over cycles of selection. (A) Bulk variants selected from each cycle were co-transfected with gRNA targeting AAVS1 and ½ GFP transposon into the reporter cell line. Quantity of plasmid was corrected by PB copy number to normalize for cloning efficiency. (B) Lentiviruses expressing bulk variants of each cycle were produced and used to infect reporter the cell line.

Figure 23. Specific target integration relative to FiCAT (hyPB R372A-K375A-D450N) of single mutants isolated from bulk variants after 4 and 5 cycles of cas9_PB library enrichment co transfected with gRNA tcr1 and ½ GFP MC transposon.

Figure 24. Programmable insertion activity of dimeric hyPB R372A-K375A-D450N fused with either SpCas9 or SaCas9 for targeted insertion in ½ GFP reporter cell line.

Figure 25. Relative comparison of the programmable insertion activity for targeted insertion in ½ GFP reporter cell line. (A) Comparison between hyPB R372A-K375A-D450N fused with SpCas9 protein (left) and hyPB R372A-K375A-D450N fused with MCP protein with SpCas9 added separately (right). (B) Comparison between 3 hyPB mutants (R372A-K375A-D450N; R202K-R275A-N347S-R372A-D450N-T560A-F594L; and R275A-N347S-R372A-D450N-T560A-F594L) fused to MCP protein with SpCas9 added separately.

Figure 26. Comparison of the programmable insertion activity for targeted insertion in ½ GFP reporter cell line. (A) Comparison between the co-expression of hyPB R372A-K375A-D450N and SpCas9 protein (left) and the fusion protein comprising hyPB R372A-K375A-D450N with SpCas9 protein (right). (B) Relative comparison between 3 hyPB mutants (R372A-K375A-D450N; R202K-R275A-N347S-R372A-D450N-T560A-F594L; and R275A-N347S-R372A-D450N-T560A-F594L) co-expressed with SpCas9.

Figure 27. Relative comparison of the programmable insertion activity for targeted insertion in ½ GFP reporter cell line with the co-expression of a first fusion protein comprising SpCas and hyPB R372A-K375A-D450N, and a second fusion protein comprising MCP protein and hyPB mutants (R372A-K375A-D450N; R202K-R275A-N347S-R372A-D450N-T560A-F594L; and R275A-N347S-R372A-D450N-T560A-F594L).

Figure 28. Relative comparison of the programmable insertion activity for targeted insertion in ½ GFP reporter cell line with the co-expression of a fusion protein comprising SpCas and hyPB R372A-K375A-D450N, and 3 hyPB mutants R372A-K375A-D450N; R202K-R275A-N347S-R372A-D450N-T560A-F594L; and R275A-N347S-R372A-D450N-T560A-F594L.

Figure 29. Comparison of the programmable insertion activity for targeted insertion in ½ GFP reporter cell line between SpCas9 fused to a dimer of hyPB R272A-K275A-D450N

(left) and SpCas9 fused to a first hyPB R272A-K275A-D450N and to a second hyPB mutant (right).

DEFINITIONS

As used herein, the singular forms "**a**", "**an**", and "**the**" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

The terms "**nucleic acid sequence**" and "**nucleotide sequence**" may be used interchangeably to refer to any molecule composed of, or comprising, monomeric nucleotides. A nucleic acid may be an oligonucleotide or a polynucleotide. A nucleotide sequence may be a DNA, RNA, or a mix thereof. A nucleotide sequence may be chemically-modified or artificial. Nucleotide sequences include peptide nucleic acids (PNA), morpholinos and locked nucleic acids (LNA), as well as glycol nucleic acids (GNA) and threose nucleic acid (TNA). Each of these sequences is distinguished from naturally-occurring DNA or RNA by changes to the backbone of the molecule. Also, phosphorothioate nucleotides may be used. Other deoxynucleotide analogs include, without limitation, methylphosphonates, phosphoramidates, phosphorodithioates, N3'P5'-phosphoramidates and oligoribonucleotide phosphorothioates and their 2'-*O*-allyl analogs and 2'-*O*-methylribonucleotide methylphosphonates which may be used in a nucleotide of the disclosure.

The term "**transgene**" refers to an exogenous nucleic acid sequence, in particular an exogenous DNA or cDNA encoding a gene product. The gene product may be an RNA, peptide or protein. In addition to the coding region for the gene product (CDS), the transgene may include or be associated with one or more operational sequences to facilitate or enhance expression, such as a promoter, enhancer(s), response element(s), reporter element(s), insulator element(s), polyadenylation signal(s) and/or other functional elements. Embodiments of the disclosure may utilize any known suitable promoter, enhancer(s), response element(s), reporter element(s), insulator element(s),

polyadenylation signal(s) and/or other functional elements, unless specified otherwise. Suitable elements and sequences will be well known to those skilled in the art.

The terms "polypeptide", "peptide", and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

The term "**binding protein**" refers to a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to one or more molecules of the same protein to form homodimers, homotrimers, etc.; and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

The terms "**Cas9**" or "**Cas9 nuclease**" refer to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems, correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs

("sgRNA" or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species.

Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self *vs.* non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art. Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski *et al.*, **2013**. (*RNA Biol.* **10(5)**:726-37), the entire content of which is incorporated herein by reference.

In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain. A nuclease-inactivated Cas9 protein can interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known in the art (see, *e.g.*, Jinek *et al.*, **2012**. *Science*. **337(6096)**:816-821; Qi et al., **2013**. *Cell*. **152(5)**:1173-83, the entire content of each being incorporated herein by reference).

The term "**zinc finger protein**" refers to a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequences within a binding domain of the zinc finger protein whose structure is stabilized through coordination of a zinc ion. The term "zinc finger protein" is often abbreviated as "ZFP".

The term "**zinc finger nuclease**" refers to an artificial restriction enzyme generated by fusing a zinc finger DNA-binding domain to a DNA-cleavage domain. Zinc finger domains can be engineered to target specific desired DNA sequences, and this enables zinc finger nucleases to target unique sequences within complex genomes. "Zinc finger nuclease" is often abbreviated as "ZFN" or "ZNP".

The term "amino acid sequence" or "polypeptide" or "protein" as used herein, refers a polymer of amino acid residues. Unless specified, a polymer of amino acid residues can be any length.

The term "exogenous" as used herein, refers to a molecule that is not naturally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. Natural presence in the cell may also be determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally functioning endogenous molecule.

By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

A "**target sequence**" or "**target nucleic acid sequence**" or "**target site**" is a sequence that defines a portion of a nucleic acid, *e.g.*, in a genome, to which a binding molecule will bind, provided sufficient conditions for binding exist. For example, the sequence 5'-GAATTC-3' is a target site for the EcoRI restriction endonuclease.

The term "**fusion**" refers to a molecule in which two or more subunit molecules are linked. In some embodiments, the link between the two is covalent; alternatively, the link between the two can be non-covalent and rely, *e.g.*, on intermolecular interactions. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules.

The term "**fusion protein**" refers to a hybrid polypeptide which comprises protein domains from at least two different proteins. For example, one protein domain may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein, thus forming an "amino-terminal fusion protein"

or a “carboxy-terminal fusion protein”, respectively. In preferred embodiments, a fusion protein is a single chain polypeptide which may be fully encoded by a nucleic acid sequence, and includes at least two protein domains directly covalently linked by peptidic bound or optionally covalently linked via a peptidic linker.

The terms "gene" or "genome" as used herein, includes a DNA region encoding a gene product, as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

The term "eukaryotic," cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (e.g., T-cells).

The term "linked" as used herein, refers to the juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components.

A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid, respectively, whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions.

The term "transfect" as used herein, refers to the introduction of nucleic acids (either DNA or RNA) into eukaryotic or prokaryotic cells or organisms.

The term “**cleavage**” refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic

or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

The term "**specificity**" refers to the ability to selectively bind a sequence which shares a degree of sequence identity to a selected sequence.

The terms "**insertion**" and "**integration**" refer to the addition of a nucleic acid sequence into a second nucleic acid sequence or into a genome or part thereof. The terms "**specific**", "**site-specific**", "**targeted**" and "**on-targeted**" in relation to insertion or integration, are used herein interchangeably to refer to the insertion of a nucleic acid into a specific site of a second nucleic acid or into a specific site of a genome or part thereof. Conversely, the terms "**random**", "**non-targeted**" and "**off-targeted**" refer to non-specific and unintended insertion of a nucleic acid into an unwanted site. The terms "**total**" or "**overall**" refer to the total number of insertions.

The term "**mutation**" refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue; and/or to a deletion or insertion of one or more residues within a nucleic acid or amino acid sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence, then the identity of the newly substituted residue. Various methods for making amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green & Sambrook, **2012** (*Molecular cloning: a laboratory manual* (4th Ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). In preferred embodiments, the term mutation in a protein refers to an amino acid substitution.

The term "**transposase**" refers to an enzyme that binds to the end of a transposon and catalyzes its movement to another part of the genome by a cut-and-paste mechanism or a replicative transposition mechanism.

The term "**modified**" refers to a protein or nucleic acid sequence that is different than a corresponding unmodified protein or nucleic acid sequence.

The term "**linker**" refers to a chemical group or a molecule linking two adjacent molecules or moieties.

The terms "vector" and "plasmid" as used herein, refer to any polynucleotide that can carry, e.g., a second polynucleotide of interest, and e.g., which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors. Particularly, the term "expression vector," as used herein, refers to any polynucleotide capable of directing the expression of a nucleic acid. In some aspects, the terms "vector" and "plasmid" are used interchangeably with the term "nucleic acid construct."

As used herein, the percent identity between two sequences is a function of the number of identical positions shared by the sequences (i. e., % identity = number of identical positions/total number of positions x 100), taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm, as described below.

The percent identity between two amino acid sequences can be determined using the algorithm of E. Meyers and W. Miller (Comput. Appl. Biosci., 4:11-17, 1988) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Alternatively, the percent identity between two amino acid sequences can be determined using the Needleman and Wunsch (J. Mol, Biol. 48:444-453, 1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

The percent identity between two nucleotide amino acid sequences may also be determined using for example algorithms such as the BLASTN program for nucleic acid

sequences using as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The terms "recombinant" or "engineered," as used herein, refer to a protein or nucleic acid sequence that has been artificially created.

The term “subject” as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal.

The terms “treatment,” “treat,” and “treating,” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms “treatment,” “treat,” and “treating” refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent, reduce the likelihood of developing, or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

DETAILED DESCRIPTION

The present invention relates to a composition comprising

(i) a first protein comprising or consisting of a site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence; or a nucleic acid construct encoding said first protein;

(ii) a second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein; and

wherein said transposase is a modified hyperactive PiggyBac, comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9.

Current genome engineering tools, including engineered zinc finger proteins (ZFPs), transcription activator like effector nucleases (TALENs), and more recently, the RNA-guided DNA nucleases such as Cas9, effect sequence-specific DNA cleavage in a genome. This programmable cleavage can result in mutation of the DNA at the cleavage site via non-homologous end joining (NHEJ) or replacement of the DNA surrounding the cleavage site via homology-directed repair (HDR).

In one embodiment, the site-specific DNA binding protein is selected from the group comprising or consisting of RNA-guided DNA nucleases, zinc finger proteins and transcription activator like effector nucleases.

In one embodiment, the site-specific DNA binding protein is selected from the group comprising or consisting of RNA-guided DNA nucleases and zinc finger proteins.

In one embodiment, the site-specific DNA binding protein is an RNA-guided nuclease.

In one embodiment, the site-specific DNA binding protein is a Cas9 protein (*e.g.*, without limitation, *Streptococcus pyogenes* Cas9 (SpCas9), *Staphylococcus aureus* Cas9 (SaCas9), or *Campylobacter jejuni* Cas9 (CjCas9); some other suitable examples will be described below), or a variant thereof (*e.g.*, nickase Cas9 (nCas9) or dead Cas9 (dCas9)), a Cas12a protein, a Cas12b protein, a Cpf1 protein, or a CasX protein, including variants and functional fragments thereof.

In one embodiment, the site-specific DNA binding protein is a Cas9 protein, including variants and functional fragments thereof.

The CRISPR-Cas9 system is a highly effective tool for inactivating or modifying genes via sequence-specific double-strand breaks (DSBs). These DSBs are recognized by the cellular DNA damage response machinery and can be repaired by endogenous DSB repair pathways. The predominant repair pathway is non-homologous end joining (NHEJ), which often results in small insertions and/or deletions that can create frameshift mutations and disrupt the function of genes. This pathway can be exploited to generate genetic knockout mutations. Alternatively, in the presence of repair templates (such as, *e.g.*, the nucleic acid construct comprising or consisting of a transgene encoding the laminin-α2 protein, functional variant or fragment thereof), the damage can be repaired seamlessly by homology-directed repair (HDR). However, despite remarkable progress, HDR-mediated genome editing to introduce precise genetic modifications is much less efficient than NHEJ-mediated gene disruption. Furthermore, large multi-kb replacements by the HDR pathways results challenging and requires selection and/or large population cell sorting. Consequently, the major applications for the HDR pathways are currently limited to the local replacement of key regions within genes, but not of large, full-length genes. As explained above, the present invention remedies this deficiency.

In one embodiment, the Cas9 protein comprises (i) an active DNA cleavage domain and (ii) a guide RNA binding domain.

Among the known Cas9 proteins, the *S. pyogenes* Cas9 protein has been widely used as a tool for genome engineering. This Cas9 protein is a large, multi-domain protein containing two distinct nuclease domains.

In one embodiment, the Cas9 protein is selected from the group comprising or consisting of the Cas9 protein from *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1) with SEQ ID NO: 19); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1) with SEQ ID NO: 20; *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1) with SEQ ID NO: 21; *Prevotella intermedia* (NCBI Ref: NC_017861.1) with SEQ ID NO: 22; *Spiroplasma taiwanense* (NCBI Ref:

NC_021846.1) with SEQ ID NO: 23; *Streptococcus iniae* (NCBI Ref: NC_021314.1) with SEQ ID NO: 24; *Belliella baltica* (NCBI Ref: NC_018010.1) with SEQ ID NO: 25; *Psychroflexus torquisi* (NCBI Ref: NC_018721.1) with SEQ ID NO: 26; *Streptococcus thermophilus* (NCBI Ref: YP_820832.1) with SEQ ID NO: 27; *Listeria innocua* (NCBI Ref: NP_472073.1) with SEQ ID NO: 28; *Campylobacter jejuni* (CjCas9) (NCBI Ref: YP_002344900.1) with SEQ ID NO: 29 (encoded by SEQ ID NO: 81); *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) with SEQ ID NO: 30; *Staphylococcus aureus* (SaCas9) with SEQ ID NO: 72 (encoded by SEQ ID NO: 77); and *Streptococcus pyogenes* (SpCas9) (NCBI Ref: NC_017053.1) with SEQ ID NO: 31.

In one embodiment, when referring herein to the wild-type Cas9 protein, said wild-type Cas9 protein corresponds to Cas9 from *Streptococcus pyogenes* (spCas9) with SEQ ID NO: 31, unless specified otherwise.

In one embodiment, the Cas9 protein may be a "Cas9 variant". A "**Cas9 variant**", as used herein, is a protein sharing homology to a Cas9 protein as described herein, and includes fragments thereof.

In one embodiment, the Cas9 variant can be at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to a wild-type Cas9 protein with SEQ ID NO: 31, or to any other Cas9 protein with SEQ ID NOs: 19-30 or 72.

In one embodiment, the Cas9 variant comprises the amino acid sequence of a Cas9 protein with one or several amino acid substitutions. For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand.

Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the substitutions D10A and H841A are known to completely inactivate the

nuclease activity of the *S. pyogenes* Cas9 protein with SEQ ID NO: 31, resulting in a dead Cas9 (dCas9) that still retains its ability to bind DNA in a sgRNA-programmed manner. In principle, when fused to another protein or domain, dCas9 can target that protein to virtually any DNA sequence simply by co-expression with an appropriate sgRNA. In one embodiment, the dCas9 protein is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 66. In one embodiment, the dCas9 protein comprises or consists of an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 71.

As to Cas9 nickase (nCas9), it is a variant of Cas9 nuclease differing by a point mutation (D10A) in the RuvC nuclease domain, which enables it to nick, but not cleave, DNA. In one embodiment, the nCas9 protein is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 65. In one embodiment, the nCas9 protein comprises or consists of an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 70. In some embodiments, the SaCas9 nickase (SanCas9) is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 80. In some embodiments, the SaCas9 nickase (SanCas9) comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 76.

In one embodiment, the Cas9 variant comprises a fragment of Cas9, such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of a wild-type Cas9 protein with SEQ ID NO: 31, or of any other Cas9 protein with SEQ ID NOs: 19-30 or 72.

In one embodiment, the Cas9 variant comprises only one of a DNA cleavage domain or a guide RNA binding domain.

In one embodiment, an exemplary Cas9 variant is humanized Cas9 (hCas9) or a variant or functional fragment thereof. As used herein, the term "**humanized Cas9**" or "**hCas9**" refers to a sequence-optimized Cas9 protein for human cells.

In one embodiment, the hCas9 protein is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 64. In one embodiment, the hCas9 protein comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 69.

In one embodiment, the site-specific DNA binding protein is a cpf1 protein. In one embodiment, the cpf1 protein is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 78. In one embodiment, the cpf1 protein comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 74.

In one embodiment, the site-specific DNA binding protein is a CasX protein. In one embodiment, the CasX is encoded by a nucleic acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 79. In one embodiment, the CasX comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to SEQ ID NO: 75.

As will be further detailed below, certain aspects of the disclosure are also directed to vectors or plasmids (*e.g.*, expression vectors, packaging vectors, etc.) comprising a nucleic acid construct encoding the site-specific DNA binding protein, in particular the RNA-guided nuclease, in particular any of the Cas9 proteins described herein; said vectors or plasmids being preferably suitable for expression in a host cell, *e.g.*, mammalian cells, yeast cells, insect cells, plant cells, fungal cells, or algal cells.

In one embodiment, the site-specific DNA binding protein is a zinc finger protein (ZFP).

Zinc finger proteins are proteins that can bind to DNA in a sequence-specific manner. ZFP are unevenly distributed in eukaryotes. ZFP have been identified that are involved in DNA recognition, RNA binding, and protein binding. Certain classifications for zinc finger proteins are based on "fold groups" in view of the overall shape of the protein backbone in the folded domain. The most common "fold groups" of zinc fingers are the C2H2 or Cys2His2-like (the "classic zinc finger"), treble clef, and zinc ribbon. Representative motifs characterizing these proteins are disclosed in Table 1 of Li & Liu, **2020** (*Int J Mol Sci*. **21(4)**:1361), which Table is herein incorporated by reference.

The ZFP can be any ZFP, variant or functional fragment thereof, that can bind to a specific genomic DNA sequence in a genome. Non-limiting examples of ZFPs include ZFPs comprising a fold group or zinc finger motif selected from C2H2, gag knuckle, treble clef, zinc ribbon, Zn2/Cys6-like, or TAZ2 domain-like, or any combination thereof. In one embodiment, the ZFP is a C2H2 zinc finger protein.

In one embodiment, the ZFP is an engineered ZFP. Engineered zinc finger arrays can be fused to a DNA cleavage domain (usually the cleavage domain of FokI) to generate zinc finger nucleases. Such zinc finger-FokI fusions have become useful reagents for manipulating genomes.

The ZFP can comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more zinc finger domains. The ZFP can comprise from 2 to 12, from 2 to 10, from 2 to 8, from 3 to 8, from 4 to 8, or from 5 to 8 zinc finger domains. In one embodiment, the ZFP comprises 6 zinc finger domains.

A common modular assembly process involves combining separate zinc fingers that can each recognize a 3-basepair DNA sequence to generate 3-finger, 4-, 5-, or 6-finger arrays that recognize target sites ranging from 9 basepairs to 18 basepairs in length. Another method uses 2-finger modules to generate zinc finger arrays with up to six individual zinc fingers.

In one embodiment, the binding domain of the ZFP can be engineered to bind to a sequence of interest. An engineered zinc finger binding domain can have improved binding specificity, compared to a naturally-occurring ZFP.

In one embodiment, exemplary nucleic acid sequences encoding the ZFP comprise or consists of SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, or SEQ ID NO: 38. In one embodiment, exemplary amino acid sequences encoded by these sequences comprise or consists of SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, or SEQ ID NO: 39.

In one embodiment, the ZFP comprises an amino acid sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% sequence identity to any one of SEQ ID NOs: 33, 35, 37 or 39.

In one embodiment, the ZFP does not have a Gal4 DNA binding domain. Gal4 binds to CGG-N_{11}-CCG, where N can be any base. This protein is a positive regulator for the gene expression of the galactose-induced genes such as *GAL1*, *GAL2*, *GAL7*, *GAL10*, and *MEL1* which code for the enzymes used to convert galactose to glucose. It recognizes a

17-base pair sequence in the upstream activating sequence (UAS-G) of these genes. Therefore, Gal4 recognizes a short and very frequent sequence in the genome, thus not being site-specific. In one embodiment, the ZFP has a Gal4 DNA binding domain engineered to be site-specific.

As will be further detailed below, certain aspects of the disclosure are directed to vectors or plasmids (*e.g.*, expression vectors, packaging vectors, etc.) comprising a nucleic acid construct encoding the site-specific DNA binding protein, in particular the ZFP described herein; said vectors or plasmids being preferably suitable for expression in a host cell, *e.g.*, mammalian cells, yeast cells, insect cells, plant cells, fungal cells, or algal cells.

According to the invention, the second protein comprises or consists of a transposase.

Transposons are chromosomal segments that can undergo transposition, *e.g.*, DNA that can be translocated as a whole in the absence of a complementary sequence in the host DNA. Transposons can be used to perform long-range DNA engineering in human cells. Common transposon systems used in mammalian cells include, without limitation, Sleeping Beauty (SB), which was reconstructed from inactive transposons, and PiggyBac (PB), isolated from the moth *Trichoplusia*. PiggyBac has higher transposition activity than SB and it can be excised scarlessly.

Native DNA transposons typically contain a single gene coding for a transposase protein, which is flanked by Inverted Terminal Repeats (ITRs) that carry transposase binding sites. During their transposition, the transposase protein recognizes these ITRs to catalyze excision and subsequent reintegration of the element elsewhere in a random manner. Moreover, some of these transposons can be adapted for use in gene therapy protocols, employing them as bi-component systems, in which a plasmid contains an expression cassette where a DNA sequence of interest, placed between the transposon ITRs, can be introduced into a host genome directed by a co-transfected plasmid containing the sequence encoding the transposase enzyme or its mRNA synthesized *in vitro*. According to the disclosure, a transposon-based system is used to efficiently mediate stable integration and persistent expression of transgenes in a cell, as therapeutic genes.

A transposase or modified transposase of the disclosure can be any transposase that can insert an exogenous nucleic acid into a specific site of a genome. Some aspects of this disclosure provide transposase fusion proteins that are designed using the methods and strategies described herein. Some embodiments of this disclosure provide nucleic acids encoding such transposases or modified transposases and/or fusion proteins comprising the same. Some embodiments of this disclosure provide plasmids or expression vectors comprising such nucleic acid constructs encoding transposases or modified transposases and/or fusion proteins comprising the same.

Non-limiting examples of transposases include Frog Prince, Sleeping Beauty, hyperactive Sleeping Beauty, PiggyBac, and hyperactive PiggyBac.

In one embodiment, the transposase is a hyperactive PiggyBac transposase. In some embodiments, the transposase is the hyperactive PiggyBac transposase corresponding to SEQ ID NO: 9 or as encoded by SEQ ID NO: 67 (referred in this disclosure also as hyPB or simply as PB).

In one embodiment, the transposase is a modified hyperactive PiggyBac transposase.

As used herein, "**modified hyperactive PiggyBac transposase**" refers to a transposase comprising one or more amino acid substitutions, typically no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid substitutions, as compared to the wild-type hyperactive PiggyBac transposase with SEQ ID NO: 9. More specifically, a modified hyperactive PiggyBac comprises (i) one or more amino acid substitutionsto increase excision activity as compared to the wild-type hyperactive PiggyBac transposase, and/or (ii) one or more amino acid substitutionsto decrease DNA binding activity as compared to the wild-type hyperactive PiggyBac transposase. In one embodiment, the modified hyperactive PiggyBac transposase comprises an amino acid sequence at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the sequence set forth in SEQ ID NO: 9.

In some embodiments, the one or more mutations to the hyperactive PiggyBac transposase do not consist of a triple mutation R372A/K375A/D450N, said position

numbers corresponding to the amino acid numbers of unmodified hyperactive PiggyBac of SEQ ID NO:9.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to increase excision activity.

In some embodiment, the modified hyperactive Piggybac comprises one or more amino acid mutations to increase excision activity selected among the amino acid mutations within the region defined by the amino acid position numbers [194-200], [214-222], [434-442] or [446-456], for example amino acid substitution at the position D198, D201, R202, M212 and/or S213; said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9.

In some embodiment, the modified hyperactive Piggybac comprises one or more amino acid mutations to increase excision activity selected among the amino acid mutations at positions 450, 560, 564, 573, 589, 592, and/or 594; said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO: 9.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to increase excision activity selected among the amino acid mutations at position of M194 and/or D450, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9, preferably the amino acid substitution selected among M194V and/or D450N.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to decrease DNA binding activity.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to decrease DNA binding activity selected among the amino acid mutations at positions 254, 275, 277, 347, 372, 375, and/or 465; said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO: 9.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to decrease DNA binding activity selected among R275, N347, R372,

K375, R376, E377, and E380, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to decrease DNA binding activity selected among R372, K375, R376, E377, and E380, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9, preferably selected among the amino acid substitutions R372A, K375A, R376A, E377A, and/or E380A.

In some embodiment, the modified hyperactive PiggyBac comprises one or more amino acid mutations to decrease DNA binding activity selected among N347, R372, and K375, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9, preferably selected among the amino acid substitutions N347S, N347A, R372A, K375A, more preferably selected among the amino acid substitutions N347S, N347A.

In some embodiment, the modified hyperactive Piggybac comprises one or more amino acid mutations to increase excision activity, as defined above; and one or more amino acid mutations to decrease DNA binding activity, as defined above.

In some embodiment, the modified hyperactive Piggybac includes at least one amino acid substitution to increase excision activity at position D450, and at least two amino acid substitutions to decrease DNA binding activity at positions N347, R372 and K375, preferably said modified transposase of hyperactive Piggybac includes the double mutations N347S and D450N or triple mutations D450N, R372A and K375A, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9. In a more preferred embodiment, the modified transposase of hyperactive Piggybac includes the double mutations N347S and D450N, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9.

In some embodiment, the modified hyperactive Piggybac as disclosed in the previous embodiments further comprises at least one mutation in the region defined by the amino

acid position numbers [158-169], for example A166S; and/or at least one mutation at position Y527, R518, K525, N463.

Typically, said modified hyperactive Piggybac comprises an amino acid sequence having at least 85%, at least 90%, at least 95% identity, or 100% identity to modified hyperactive Piggybac of SEQ ID NO:1.

In some embodiments, said modified hyperactive Piggybac is a variant of the hyperactive Piggybac of SEQ ID NO:9 with one or more amino acid substitutions, typically with no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions as compared to SEQ ID NO:9.

In some embodiments, said modified hyperactive Piggybac further comprises one or more of the following amino acid mutations at positions 34, 43, 117, 202, 230, 245, 268, 275, 277, 287, 290, 315, 325, 341, 346, 347, 350, 351, 356, 357, 388, 409, 411, 412, 432, 447, 460, 461, 465, 517, 560, 564, 571, 573, 576, 586, 587, 589, 592, and/or 594, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In some embodiments, said modified hyperactive PiggyBac comprises the following mutations or combination of mutations: V34M, T43I, Y177H, R202K, S230N, R245A, D268N, K287A, K290A, K287A/K290A, R315A, G325A, R341A, D346N, N347A, N347S, T350A, S351E, S351P, S351A, K356E, N357A, R388A, K409A, A411T, K412A, K432A, D447A, D447N, D450N, R460A, K461A, W465A, S517A, T560A, S564P, S571N, S573A, K576A, H586A, I587A, M589V, S592G, or F594L, D450N/R372A/K375A, R275A/R277A, K409A/K412A, R460A/K461A, R275A/R277A/N347S/K375A/T560A/S573A/M589V/S592G and R245A/R275A/R277A/R372A/W465A.

In some embodiments, said modified hyperactive PiggyBac comprises the following amino acid substitution or combination of amino acid substitutions: R372A/K375A/D450N, R372A/K375A/R376A/D450N, K375A/R376A/E377A/E380A/D450N, R372A/K375A/R376A/E377A/E380A/D450N, M194V, M194V/R372A/K375A,

S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L,R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, T43I/R372A/K375A/A411T/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

Very preferred modified hyperactive PiggyBac transposases for use according to the present disclosure include modified hyperactive PiggyBac comprising the following combination of amino acid substitutions: R372A/K375A/D450N, S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L,R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, T43I/R372A/K375A/A411T/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G and

R275A/325A/R372A/T560A, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In some embodiments, said modified hyperactive PiggyBac comprises the following amino acid substitution or combination of amino acid substitutions: R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G, , the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In a more preferred embodiment, said modified hyperactive PiggyBac comprising the following combination of amino acid substitutions: N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G, the position number

corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 1-8, 10-18 and 135-149.

In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 1-8 and 10-18.

In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 90-99.

In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 135-149. In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 135-140. In some embodiments, said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 141-149.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are involved in the conserved catalytic triad, e.g., at amino acid 268 and/or 346 (e.g., D268N and/or D346N) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 11.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are critical for excision, e.g., at amino acid 287, 287/290 and/or 460/461 (e.g., K287A, K287A/K290A, and/or R460A/K461A) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 12.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are involved in target joining, e.g., at amino acid 351, 356, and/or 379 (e.g., S351E, S351P, S351A, and/or K356E) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 13.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are critical for integration, e.g., at amino acid 560, 564, 571, 573,

589, 592, and/or 594 (e.g., T560A, S564P, S571N, S573A, M589V, S592G, and/or F594L) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 14.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are involved in alignment, e.g., at amino acid 325, 347, 350, 357 and/or 465 (e.g., G325A, N347A, N347S, T350A and/or W465A) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 15.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are well conserved, e.g., at amino acid 576 and/or 587 (e.g., K576A and/or I587A) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 16.

In some embodiments, the modified transposase can comprise one or more mutations relative to hyPB that are involved in Zn2+ binding, e.g., 586 (e.g., H586A) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 17.

In some embodiments, the programmable transposase can comprise one or more mutations relative to hyPB that are involved in integration e.g., 315, 341, 372, and/or 375 (e.g., R315A, R341A, R372A, and/or K375A) corresponding to the amino acid numbering of SEQ ID NO: 9 or SEQ ID NO: 18.

In some embodiments, the modified hyperactive PiggyBac comprises an amino acid sequence at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the sequence set forth in SEQ ID NO: 9. In some embodiments, the modified hyperactive PiggyBac is selected for its high specificity of DNA integration into a genome compared to hyperactive PiggyBac. In some embodiments, the modified hyperactive PiggyBac comprises an amino acid sequence having one or more of the modifications disclosed herein relative to SEQ ID NO: 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18, and retains at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18, respectively.

In some embodiments, the hyperactive PiggyBac transposase is encoded by a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 67. In some embodiments, the SB100 transposase is encoded by a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 68.

In some embodiments, the SB100 transposase comprises an amino acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 73.

In some embodiments, the modified transposase is a modified Sleeping Beauty transposase comprising one or more mutations. In some embodiments, the one or more mutations in Hyper Active Sleeping Beauty Transposase or SB100 corresponds to: L25F, R36A, I42K, G59D, I212K, N245S, K252A and Q271L of SEQ ID NO: 9 or SEQ ID NO: 73.

In certain embodiments, the modified transposase is not a Himar1C9 mutant.

Certain aspects of the disclosure are directed to a vector or a plasmid (e.g., an expression vector or a packaging vector) comprising a nucleic acid construct comprising a transposase or a modified transposase of the disclosure suitable for expression in a host cell, e.g., mammalian cells, yeast cells, insect cells, plant cells, fungal cells, or algal cells. In some embodiments, the modified transposase is expressed as a fusion protein with a Cas9. In some embodiments, the modified transposase is co-expressed with a Cas9 from separate vectors, but delivered to the same cell. In some embodiments, the modified transposase or the fusion protein comprising the same is packaged in a lentivirus particle for delivery to a cell.

As shown in the examples, newly developed hyperactive PiggyBac transposase mutations library have been used to identify modified hyperactive PiggyBac which perform specific targeted transpositions. Modified hyperactive PiggyBac with positive targeted transposition were identified using such library.

In some embodiments, the modified hyperactive PiggyBac transposase can comprise a mutation of one or more of amino acids selected from amino acid: 245, 275, 277, 325, 347, 351, 372, 375, 388, 450, 465, 560, 564, 573, 589, 592, 594 corresponding to the amino acid numbering of SEQ ID NO: 9.

In some embodiments, the modified hyperactive PiggyBac transposase mutation can comprise one or more of the amino acid modifications selected from: R245A, R275A, R277A, R275A/R277A, G325A, N347A, N347S, S351E, S351P, S351A, R372A, K375A, R388A, D450N, W465A, T560A, S564P, S573A, M589V, S592G, or F594L corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modification D450N corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase correspond to SEQ ID NO:1 and comprises the amino acid modifications R372A, K375A and D450.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modifications R245A and D450, corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modifications R245A, G325A, and S573P, corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modifications R245A, G325A, D450 and S573P, corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modification N347S or N347A, corresponding to the amino acid numbering of SEQ ID NO: 9.

In an embodiment, the modified hyperactive PiggyBac transposase comprises the amino acid modifications N347S and D450N, corresponding to the amino acid numbering of SEQ ID NO: 9.

In another, the modified hyperactive PiggyBac transposase comprises the amino acid modifications N347A and D450N, corresponding to the amino acid numbering of SEQ ID NO: 9. In one embodiment, this modified hyperactive PiggyBac transposase comprises the amino acid sequence of SEQ ID NO: 137.

As said before, herein provided are modified hyperactive PiggyBac transposases which can be fused to the elements disclosed herein but can also be used alone or in combination with different elements. Said transposases have been generated by the inventors. Thus, modified hyperactive PiggyBac transposases are provided which comprises the amino acid sequence SEQ ID NO: 9, wherein:

- amino acid at position 34 is V or M,
- amino acid at position 43 is T or I,
- amino acid at position 177 is Y or H,
- amino acid at position 202 is R or K,
- amino acid at position 230 is S or N,
- amino acid at position 245 is A,
- amino acid at position 268 is D or N,
- amino acid at position 277 is R or A,
- amino acid at position 275 is R or A,
- amino acid at position 277 is R or A,
- amino acid at position 325 is A or G,
- amino acid at position 347 is S, or A,
- amino acid at position 351 is E, P or A,
- amino acid at position 372 is R or A,
- amino acid at position 375 is K or A,
- amino acid at position 388 is R or A,
- amino acid at position 409 is K or A,
- amino acid at position 411 is A or T,

- amino acid at position 412 is K or A,
- amino acid at position 450 is D or N,
- amino acid at position 460 is R or A,
- amino acid at position 465 is W or A,
- amino acid at position 517 is S or A,
- amino acid at position 560 is T or A,
- amino acid at position 564 is P or S,
- amino acid at position 571 is S or N,
- amino acid at position 573 is S or A,
- amino acid at position 576 is K or A,
- amino acid at position 586 is H or A,
- amino acid at position 587 is I or A,
- amino acid at position 589 is M or V,
- amino acid at position 592 is G or S, and/or,
- amino acid at position 594 is L or F.

The present disclosure also relates to the modified hyperactive PiggyBac transposases provided herein for use as medicaments, particularly in gene therapy, ex vivo or in vivo.

In one embodiment, the first protein comprising or consisting of the site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence (as described above), and the second protein comprising or consisting of a transposase (as described above), are fused together to form a fusion protein, either directly or indirectly via a linker.

Any embodiments relating to the site-specific DNA binding protein on one hand, and to the transposase on the other hand, apply *mutatis mutandis* in the case of the fusion protein described herein.

Hence, in one embodiment, the fusion protein comprises or consists of:

(i) a first protein comprising or consisting of an RNA-guided DNA nuclease, a zinc finger protein or a transcription activator like effector nuclease, as described above, and

(ii) a second protein comprising or consisting of a transposase, said transposase being a modified hyperactive PiggyBac comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In one embodiment, the fusion protein comprises or consists of:

(i) a first protein comprising or consisting of an RNA-guided DNA nuclease or zinc finger protein, as described above, and

(ii) a second protein comprising or consisting of a transposase, said transposase being a modified hyperactive PiggyBac comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In one embodiment, the fusion protein comprises or consists of:

(i) a first protein comprising or consisting of an RNA-guided DNA nuclease, as described above, and

(ii) a second protein comprising or consisting of a transposase, said transposase being a modified hyperactive PiggyBac comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In one embodiment, the fusion protein comprises or consists of:

(i) a first protein comprising or consisting of a Cas9 protein or a variant thereof, as described above, and

(ii) a second protein comprising or consisting of a transposase, said transposase being a modified hyperactive PiggyBac comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In one embodiment, the first protein and the second protein can be oriented in the fusion protein in either order.

In one embodiment, the fusion protein comprises or consists of the first protein fused at the C-terminal end of the second protein, either directly or indirectly via a linker. In other words, the fusion protein comprises or consists of, from N- to C-terminal: (i) the second protein (*i.e.*, the transposase); (ii) optionally, a linker; and (iii) the first protein (*i.e.*, the site-specific DNA binding protein, preferably the RNA-guided DNA nuclease; more preferably the Cas9 protein or variant thereof).

In one embodiment, the fusion protein comprises or consists of the first protein fused at the N-terminal end of the second protein, either directly or indirectly via a linker. In other words, the fusion protein comprises or consists of, from N- to C-terminal: (i) the first protein (*i.e.*, the site-specific DNA binding protein, preferably the RNA-guided DNA nuclease; more preferably the Cas9 protein or variant thereof); (ii) optionally, a linker; and (iii) the second protein (*i.e.*, the transposase).

In one embodiment, the fusion protein comprises a linker.

Suitable examples of linkers include peptidic linkers, between the first protein and the second protein (in any order).

In one embodiment, the peptidic linker is selected from the group comprising or consisting of $(GGS)_n$, $(GGGGS)_n$ with SEQ ID NO: 133, $(G)_n$, $(EAAAK)_n$ with SEQ ID NO: 134, XTEN linkers, and $(XP)_n$ motif, and combinations of any of any of these, wherein n is independently an integer between 1 and 50.

In one embodiment, the linker is 12- to 24-amino acid long, or is encoded by a nucleic acid sequence that is 36- to 72- nucleotide long.

In one embodiment, the linker is a XTEN linker or a $(GGS)_n$linker.

In one embodiment, the linker is selected among the linkers shown in **Table 1**.

Table 1: Linkers

Linker	**Nucleic acid sequence (SEQ ID NO)**	**Amino acid sequence (SEQ ID NO)**
GGSx3	ggtggatctggcggtggatctggtggcggt (SEQ ID NO: 48)	GGSGGGSGGG (SEQ ID NO: 49)
GGS4x	ggagggagtggtgggtccggtggtagtggcggatcc (SEQ ID NO: 50)	GGSGGSGGSGGS (SEQ ID NO: 51)
GGS5x	ggaggctccggtgggtctggtgggagcggtggtagtgg cggatcc (SEQ ID NO: 52)	GGSGGSGGSGGSGGS (SEQ ID NO: 53)

GGS6x	ggaggcagtggtgggagcggtggttccgggggtagtg gtggttccgggggatcc (SEQ ID NO: 54)	GGSGGSGGSGGSGGSGGS (SEQ ID NO: 55)
GGS7x	ggaggttctggaggctccggtgggtccgggggaagtg gggggtcaggcggatcaggaggatcc (SEQ ID NO: 56)	GGSGGSGGSGGSGGSGGS GGS (SEQ ID NO: 57)
GGS8x	ggaggtagcggaggttccggagggagcggcgggagt gggggaagcgggggaagtggaggatccgggggagg atcc (SEQ ID NO: 58)	GGSGGSGGSGGSGGSGGS GGS (SEQ ID NO: 59)
Linker XTEN	tccggtagcgaaacaccggggacttcagaatcggccac cccggagtct (SEQ ID NO: 60)	SGSETPGTSESATPES (SEQ ID NO: 61)
Linker B	ggaagcgccggtagtgcggctgggtctggcgagttc (SEQ ID NO: 62)	GSAGSAAGSGEF (SEQ ID NO: 63)

In one embodiment, the linker comprises an amino acid sequence selected from the group comprising or consisting of SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 61, SEQ ID NO: 63, or any combination thereof; respectively encoded by the exemplary nucleic acid sequence of SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, SEQ ID NO: 60, SEQ ID NO: 62.

In one embodiment, the linker comprises or consists of the amino acid sequence of SEQ ID NO: 49; encoded by the exemplary nucleic acid sequence of SEQ ID NO: 48.

Also provided herein are fusion proteins obtained from the expression of any of the nucleic acid constructs provided in this disclosure.

In one embodiment, the fusion protein is a triple fusion protein.

Such triple fusion protein can comprise or consist of:

- one first protein (*i.e.*, one site-specific DNA binding protein) and two second protein (*i.e.*, two transposases); or

- two first protein (*i.e.*, two site-specific DNA binding proteins) and one second protein (*i.e.*, one transposase).

In one embodiment, the triple fusion comprises or consists of one first protein (*i.e.*, one site-specific DNA binding protein) and two second protein (*i.e.*, two transposases), and the triple fusion comprises from N- to C-terminal:

- (i) the site-specific DNA binding protein, (ii) a first transposase; (iii) a second transposase; or
- (i) a first transposase; (ii) the site-specific DNA binding protein, (iii) a second transposase; or
- (i) a first transposase; (ii) a second transposase, (iii) the site-specific DNA binding protein.

In one embodiment, the first and second transposases are identical. In one embodiment, the first and second transposases are different. For example, the first transposase can be a hyperactive PiggyBac transposase and the second transposase can be a modified hyperactive PiggyBac transposase, chosen among any of the modified hyperactive PiggyBac transposases described herein. Alternatively, both the first and second transposases can be modified hyperactive PiggyBac transposases, but each bearing a different substitution or different combination of substitutions as described herein.

In one embodiment, the first and second transposases are capable of forming a functional dimer.

In one embodiment, the triple fusion comprises or consists of two first protein (*i.e.*, two site-specific DNA binding proteins) and one second protein (*i.e.*, one transposase), and the triple fusion comprises from N- to C-terminal:

- (i) a first site-specific DNA binding protein, (ii) a second site-specific DNA binding protein; (iii) the transposase; or
- (i) a first site-specific DNA binding protein; (ii) the transposase, (iii) a second site-specific DNA binding protein; or
- (i) the transposase; (ii) a first site-specific DNA binding protein, (iii) a second site-specific DNA binding protein.

In one embodiment, the first and second site-specific DNA binding proteins are identical. In one embodiment, the first and second site-specific DNA binding proteins are different. For example, the first site-specific DNA binding protein can be a Cas9 protein and the second site-specific DNA binding protein can be a variant of a Cas9 protein, chosen among any of the Cas9 protein variants described herein. Alternatively, both the first and second site-specific DNA binding proteins can be Cas9 protein variants, but each being a different variant.

In one embodiment, the triple fusion protein optionally comprises a linker between two of its proteins or between the three proteins.

Also disclosed herein is a fusion protein comprising:

(i) the second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein, as described above, and

(ii) an RNA-binding protein capable of binding to at least one specific RNA sequence; or a nucleic acid construct encoding said RNA-binding protein.

In one embodiment, the fusion protein comprises a linker, as described above.

In one embodiment, the second protein comprises or consists of a transposase, said transposase being a hyperactive PiggyBac with SEQ ID NO: 9. In one embodiment, the second protein comprises or consists of a transposase, said transposase being a modified hyperactive PiggyBac comprising one or more amino acid mutations as compared to the hyperactive PiggyBac with SEQ ID NO: 9. In particular, the modified hyperactive PiggyBac can be any of those disclosed herein.

In one embodiment, the transposase/RNA-binding protein fusion can be further fused to the first protein comprising or consisting of the site-specific DNA binding protein, as described above.

In some embodiments, the RNA-binding protein is a MS2 bacteriophage coat protein (MCP) or a fragment thereof.

In some embodiments, the MCP has at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 151 (encoded, *e.g.*, by the nucleic acid sequence with SEQ ID NO: 150).

In some embodiments, the RNA-binding protein is capable of binding to at least one specific RNA sequence, said RNA sequence comprising a tetraloop. The term "**tetraloop**" is used interchangeably with the terms "stem loop" and "hairpin loop".

In some embodiments, the at least one tetraloop is a MS2 RNA tetraloop-binding sequence.

In some embodiments, the tetraloop is comprised within a guide RNA (gRNA). In certain embodiments, the gRNA is in a complex with a Cas9 protein, as described above.

In some embodiments, the gRNA comprises at least one MS2 RNA tetraloop-binding sequence. In some embodiments, the gRNA comprises more than one MS2 RNA tetraloop-binding sequences.

In some embodiments, the gRNA comprising the at least one MS2 RNA tetraloop-binding sequence has at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity with SEQ ID NO: 153 (encoded, *e.g.*, by the DNA sequence with SEQ ID NO: 152).

In some embodiments, the MCP in the fusion protein binds non-covalently to at least one MS2 RNA tetraloop-binding sequence comprised in a gRNA itself non-covalently bound to a Cas9 protein; in particular, the binding of the fusion protein to the Cas9/gRNA complex directs the excision activity of the modified hyperactive PiggyBac transposase towards the site specifically recognized by the Cas9/gRNA complex.

As will be further detailed below, certain aspects of the disclosure are also directed to vectors or plasmids (*e.g.*, expression vectors, packaging vectors, etc.) comprising a nucleic acid construct encoding the fusion protein described herein; said vectors or plasmids being preferably suitable for expression in a host cell, *e.g.*, mammalian cells, yeast cells, insect cells, plant cells, fungal cells, or algal cells.

According to the invention, the composition can comprise the first protein and/or the second protein (or the fusion protein comprising both), either as proteins, as described above; or as nucleic acid constructs encoding these proteins.

Targeted editing of nucleic acid sequences, *e.g.*, the introduction of a specific modification (*e.g.*, insertion of an exogenous nucleic acid) into genomic DNA, is a promising approach for treating human genetic diseases. To this end, the inventors aimed to provide improved nucleic acid constructs for use in genomic editing that are highly efficient at installing a desired modification; minimal off-target activity; and the ability to be programmed to edit precisely a site within the human genome.

Certain aspects of the present application are thus directed to a nucleic acid construct for use in improving site-specific insertion of an exogenous nucleic acid, e.g., a gene of interest (GOI), into a genome. In some embodiments, the GOI is a therapeutic gene, e.g., a gene that encodes a therapeutic protein. Examples of a therapeutic genes of interest include *CFTR* gene (Cystic fibrosis transmembrane conductance regulator) to treat Cystic Fibrosis disease; *SMN1* gene (Survival motor neuron 1) to treat Spinal muscular atrophy (SMA); *LRP5* gene (LDL receptor related protein 5) variant G171V to prevent osteoporosis and bone fractures; and *APP* gene (amyloid beta precursor protein) variant A673T to reduce Alzheimer's predisposition.

In some embodiments, the exogenous nucleic acid for insertion (e.g., the GOI) can be up to about 10 kb, up to about 15 kb, up to about 20kb in length, up to about 25kb in length, up to about 30kb in length, up to about 35kb in length, or up to about 40kb in length.

In some embodiments, the exogenous nucleic acid for insertion can be up to 10 kb, up to 15 kb, up to 20kb in length, up to 25kb in length, up to 30kb in length, up to 35kb in length, or up to 40kb in length, e.g., about 1 kb to about 40 kb, about 1 kb to about 39 kb, about 1 to about 38 kb, about 1 kb to about 37 kb, about 1 kb to about 36 kb, or about 1 kb to about 35 kb, for example and more preferably between 5 and 25 kb, typically between 8 and 20 kb.

In one embodiment, the composition of the invention comprises or consists of:

a. a nucleic acid construct encoding the first protein described above, comprising or consisting of a site-specific DNA binding protein described above;

b. a nucleic acid construct encoding a second protein, comprising or consisting of a transposase being a modified hyperactive PiggyBac, comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In another embodiment, the composition of the invention comprises or consists of a nucleic acid construct encoding the fusion protein described above, comprising or consisting of (i) a first protein comprising or consisting of a site-specific DNA binding protein, and (ii) a second protein comprising or consisting of a transposase being a modified hyperactive PiggyBac, comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9, as described above.

In one embodiment, the nucleic acid construct encoding the fusion protein further comprises a nucleic acid sequence encoding a linker between the first and the second protein, as described above; or in the case of a triple fusion protein, between two of its proteins or between the three proteins.

According to the disclosure, the first and second proteins, or the fusion protein comprising or consisting of said first and second proteins, enable and/or promote site-specific insertion of an exogenous nucleic acid.

Some embodiments are directed to a plasmid or a vector (such as, *e.g.*, an expression vector) comprising either:

- a nucleic acid construct encoding the first protein; or
- a nucleic acid construct encoding the second protein; or
- a nucleic acid construct encoding the first protein and a nucleic acid construct encoding the second protein; or
- a nucleic acid construct encoding the fusion protein or triple fusion protein.

In some embodiments, the plasmid is a packaging plasmid. In some embodiments, the plasmid further comprises a polynucleotide encoding capsid proteins, *e.g.*, gag and pol. In some embodiments, the plasmid is combined with a second plasmid comprising a

polynucleotide that encode proteins for a viral envelope (envelope plasmid); and a third plasmid comprising a nucleic acid construct comprising the exogenous nucleic acid transgene, wherein when the combination is introduced into a production cell line (*e.g.*, eukaryotic cells, prokaryotic cells and/or cell lines), a virus particle comprising the nucleic acid constructs encoding the exogenous nucleic acid transgene and the nucleic acid construct encoding either of the first protein, second protein, both first and second proteins or fusion protein, is produced.

In some embodiments, the plasmid is combined with a second plasmid comprising a polynucleotide encoding capsid proteins, *e.g.*, gag and pol (a packaging plasmid, wherein the packaging plasmid lacks a functional integrase), a third plasmid comprising a polynucleotide that encodes proteins for a viral envelope (envelope plasmid) and a fourth plasmid comprising a nucleic acid construct comprising the exogenous nucleic acid transgene, wherein when the combination is introduced into a production cell line (*e.g.*, eukaryotic and prokaryotic cells and/or cell lines), a virus particle comprising the nucleic acid constructs comprising the exogenous nucleic acid transgene and the nucleic acid construct encoding either of the first protein, second protein, both first and second proteins or the fusion protein, is produced.

In one embodiment, the first protein, second protein, both first and second proteins or fusion protein, and/or the exogenous nucleic acid transgene, are delivered to a cell using a lentivirus particle.

In one embodiment, the nucleic acid construct comprises a first polynucleotide sequence encoding the first protein comprising or consisting of site-specific DNA binding protein engineered to bind a target nucleic acid sequence, a second polynucleotide sequence encoding the second protein comprising or consisting of a transposase that enables insertion of the exogenous nucleic acid transgene into the genome, and optionally, a third polynucleotide sequence comprising a nucleic acid sequence encoding a linker between the first and second polynucleotides. In some embodiments, the first protein is a zinc finger protein or a Cas9 protein or variant thereof, as described above; and/or the second protein is a modified hyperactive PiggyBac transposase, as described above.

Examples of suitable linkers to produce a fusion protein have been described hereabove.

In some embodiments, a linker is not needed because the first protein is expressed from a separate plasmid from the second protein.

In one embodiment, instead of using a linker, the first and/or the second polynucleotide sequences comprise nucleic acids encoding the first and second protein, respectively, and further comprise additional nucleotides in at least one of their ends that make the function of linker.

In one embodiment, the nucleic acid construct is in DNA or RNA form.

Also provided herein, are vectors comprising any of the nucleic acid constructs provided in this disclosure. Particularly, the vectors are suitable for expression in mammalian cells, yeast cells, insect cells, plant cells, fungal cells, or algal cells. Also provided herein, are host cells comprising any of the nucleic acid constructs or vectors provided in this disclosure.

In some embodiments, the nucleic acid construct of the disclosure is expressed in a host cell. Suitable host cells include but not limited to eukaryotic and prokaryotic cells and/or cell lines. Non-limiting examples of such host cells or cell lines generated from such cells include COS, CHO (e.g., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (e.g., HEK293-F, HEK293-H, HEK293-T), and perC6 cells as well as insect cells such as *Spodoptera fugiperda* (Sf), or fungal cells such as *Saccharomyces*, *Pichia* and *Schizosaccharomyces*.

In some embodiments, the host cell is from a microorganism. Microorganisms which are useful for certain methods disclosed herein include, for example, bacteria (e.g., *E. coli*), yeast (e.g., *Saccharomyces cerevisiae*), and plants. The host cell can be prokaryotic or eukaryotic. In some embodiments, the host cell is eukaryotic. Suitable eukaryotic host cells include, but are not limited to, yeast cells, insect cells, plant cells, fungal cells, and algal cells.

In some embodiments, the host cell is a competent host cell. In some embodiments, the host cell is naturally competent. In some embodiments, the host cells are made competent, e.g., by a process that uses calcium chloride and heat shock. The cells used can be any cell competent, particularly eukaryotic cells, in particular mammalian, e.g. human or animal. They can be somatic or embryonic stem or differentiated. In some aspects, the cells include 293T cells, fibroblast cells, hepatocytes, muscle cells (skeletal, cardiac, smooth, blood vessel, etc.), nerve cells (neurons, glial cells, astrocytes) of epithelial cells, renal, ocular etc. It may also include, insect, plant cells, yeast, or prokaryotic cells. Additionally, primary cells may be isolated and used *ex vivo* for reintroduction into the subject to be treated following treatment with the nucleases (e.g. ZFNs or TALENs) or nuclease systems (e.g. CRISPR/Cas). Suitable primary cells include peripheral blood mononuclear cells (PBMC), and other blood cell subsets such as, but not limited to, T-lymphocytes such as CD4+ T cells or CD8+ T cells. Suitable cells also include stem cells such as, by way of example, embryonic stem cells, induced pluripotent stem cells, hematopoietic stem cells (CD34+), neuronal stem cells and mesenchymal stem cells.

In some embodiments, the host cell is transfected with a plasmid comprising a nucleic acid construct disclosed herein. In some embodiments, the plasmid comprising the nucleic acid construct is a packaging plasmid. In some embodiments, the plasmid comprising the nucleic acid construct further comprises a polynucleotide encoding capsid proteins, e.g., gag and pol. In some embodiments, the host cell is transfected with (i) the plasmid comprising the nucleic acid construct is combined in the host cell with (ii) a plasmid comprising a polynucleotide that encode proteins for a viral envelope (envelope plasmid); and (iii) a plasmid comprising an exogenous nucleic acid sequence (e.g., a GOI), wherein a virus particle comprising the exogenous nucleic acid, e.g., GOI, and the first and second proteins (either separately or as part of the fusion protein described above), is produced.

In some embodiments, the host cell is transfected with (i) the plasmid comprising the nucleic acid construct is combined with (ii) a plasmid comprising the nucleic acid construct further comprises a polynucleotide encoding capsid proteins, e.g., gag and pol (a packaging plasmid, wherein the packaging plasmid lacks a functional integrase); (iii)

a plasmid comprising a polynucleotide that encode proteins for a viral envelope (envelope plasmid) and (iv) a plasmid comprising an exogenous nucleic acid sequence (e.g., a GOI), wherein a virus particle comprising the exogenous nucleic acid, e.g., GOI, and the first and second proteins (either separately or as part of the fusion protein described above), is produced.

In further embodiments, a vector, e.g., a lentiviral vector according to the disclosure, can be used for delivering the first and second proteins (either separately or as part of the fusion protein described above) encoded by a nucleic acid construct of the disclosure and an exogenous nucleic acid to an organism, e.g., a mammal, and more particularly to a mammalian target cell of interest. The lentiviral vectors comprising the first and second proteins (either separately or as part of the fusion protein described above) are able to transduce various cell types such as, for example, liver cells (e.g. hepatocytes), muscle cells, brain cells, kidney cells, retinal cells, and hematopoietic cells. In some embodiments, the target cells of the present disclosure are “non-dividing” cells. These cells include cells such as neuronal cells that do not normally divide. However, it is not intended that the present disclosure be limited to non-dividing cells (including, but not limited to muscle cells, white blood cells, spleen cells, liver cells, eye cells, epithelial cells, etc.).

In certain embodiments, a packaged first and second proteins (either separately or as part of the fusion protein described above) is administered to an organism, e.g., for gene editing of the organism’s DNA. In some embodiments, the organism is a human. In some embodiments, the organism is a non-human mammal. In some embodiments, the organism is a non-human primate. In some embodiments, the organism is a rodent. In some embodiments, the organism is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the organism is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the organism is a research animal. In some embodiments, the organism is genetically engineered, e.g., a genetically engineered non-human subject. The organism may be of either sex and at any stage of development.

Methods for inserting a nucleic acid, for example exogenous nucleic acid, into a genome have been described. See, e.g., Yusa et al. PNAS 4(108):1531-1536 (2011); Feng et al. Nuc. Acid Res. 4(38):1204-1216 (2009); Kettlun et al. Amer. Soc. Gene and Cell Ther. 9(19):1636-1644 (2011); Skipper et al. 20(92):1-23 (2013); Li et al. PNAS 25:E2279-E2287 (2013); Mátés et al. Nature Genetics 41(6):753-761 (2009); Mali et al. Nat. Methods 10(10):957-963; Vargas et al. J. Trans. Med. 14(288):1-15 (2016); Gersbach et al. Acc. Chem. Res. 47:2309-2318 (2014); Chandrasegaran et al. Cell Gene Ther. Ins. 3(1):33-41 (2017); Wilson et al. 649:353-363 (2010); Zhao Zhang, et al. Mol Ther Nucleic Acids. 9: 230–241 (2017); Naldini L. EMBO Mol Med. 11(3) (2019); and Naldini L, et al. Hum Gene Ther. 27(10):727-728 (2016), each of which is incorporated herein by reference.

The present disclosure provides a nucleic acid construct encoding the first and second proteins (either separately or as part of the fusion protein described above), for insertion of a nucleic acid (typically exogenous nucleic acid) into a specific site of a genome. The present invention also provides the first and second proteins (either separately or as part of the fusion protein described above), for insertion of exogenous nucleic acid into a specific site of the genome. In some embodiments the exogenous nucleic acid for insertion can be up to up to 5 kb in length, up to 10 kb in length, up to 15 kb in length, 20 kb in length, up to 25kb in length, up to 30kb in length, up to 35 kb in length, or up to 40 kb in length, and in particular for long nucleic acids, for example between 5kb and 25kb, typically between 8kb and 20kb.

In another embodiment, methods for site-specific nucleic acid insertion into the genome are provided.

Hence, the present disclosure relates to a method for site specific integration of an exogenous nucleic acid sequence into the genome of a cell, the method comprising delivering to the cell, a composition comprising

(i) the first and second proteins (either separately or as part of the fusion protein described above)as disclosed herein, or a nucleic acid construct as disclosed herein,

(ii) an exogenous nucleic acid to be integrated into the genome of the cell, and

(iii)a guide RNA for determining site-specific integration of said exogenous nucleic acid into the genome of the cell.

wherein binding of said the first and second proteins (either separately or as part of the fusion protein described above) to the specific genomic DNA sequence in the genome of the cell results in cleavage of the genome and site-specific integration of said exogenous nucleic acid sequence into the genome of the cell as determined by the guide RNA.

In specific embodiments of said method, said exogenous nucleic acid is a nucleic acid fragment of a size of at least 5kb, at least 6kb, at least 7kb, at least 8 kb, at least 9kb, typically comprised between 5 and 25 kb, preferably between 8 and 20 kb.

In specific embodiments of said method, said exogenous nucleic acid is a therapeutic transgene to be inserted in a genome of a subject in need thereof to correct the deficiency of a genetic disorder.

In specific embodiments of said method, said composition is delivered in vitro or ex vivo, typically in a mammalian cell, preferably a human cell, and more preferably in a human cell which have been obtained from a human subject suffering from a genetic disorder.

In specific embodiments of said method, said composition is delivered in vivo into a mammal, for example a human subject in need thereof, typically for therapeutic treatment of a genetic disorder.

In some embodiments, the methods comprise contacting a target DNA with any of the fusion proteins comprising a Cas9 and a transposase described herein. For example, in some embodiments, the method comprises contacting a DNA with a fusion protein that comprises two linked polypeptides: (i) a Cas9; and (ii) a transposase, wherein the active Cas9 binds a gRNA that hybridizes to a region of the DNA, e.g., a genomic DNA.

In some embodiments, the methods comprise contacting a target DNA with any of the fusion proteins comprising a ZFP and an integrase described herein. For example, in some embodiments, the method comprises contacting a DNA with a fusion protein that comprises two linked polypeptides: (i) ZFP; and (ii) an integrase, wherein the active ZFP hybridizes to a region of the DNA, e.g., a genomic DNA.

In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) are delivered to an organism and/or a cell comprising the target DNA, e.g., genomic DNA, using a viral vector, e.g., a lentiviral particle.

Methods for lentiviral packaging have been described. See, Grandchamp et al. 9(6):1-13 (2014); Voelkel et al. 107(17):7805-7810 (2010); Tan et al. 80(4)1939-1948; Li et al. 9(8):1-9 (2014); Mátés et al. Nature Genetics 41(6):753-761 (2009); and Robert H Kutner1, et al. NATURE PROTOCOLS 4(4):495 (2009), each of which is incorporated herein by reference.

Typically, lentiviral delivery systems use a split system with different lentiviral genes on separate plasmids being used to produce a complete virus that does not contain the genetic components needed to cause the viral disease. For example, one plasmid (an envelope plasmid) can encode the proteins for the viral envelope (env); another plasmid (a packaging plasmid) can encode capsid proteins (e.g., gag and pol) and the enzymes like reverse transcriptase and/or integrase; and a further plasmid comprising the gene of interest (GOI) flanked by long-terminal repeats (for genome integration) and a psi-sequence (which displays a signal to package the gene into the virus) (a transfer plasmid). If these plasmids are simultaneously introduced into a cell, viruses will be produced containing the gene of interest without the viral genes that are needed to cause disease.

In certain aspects of the disclosure, the lentiviral vector (or particle) of the invention is obtainable by a split system, e.g., a transcomplementation system (vector/packaging system), by transfecting in vitro a permissive cell (such as 293T cells) with a plasmid containing certain components of the lentiviral vector genome, and at least one other plasmid providing, in trans, the gag, pol and env sequences encoding the polypeptides GAG, POL and the envelope protein(s), or for a portion of these polypeptides sufficient to enable formation of retroviral particles.

As an example, host cells are transfected with a) packaging plasmid, comprising a lentiviral gag and pol sequence, b) a second plasmid (envelope expression plasmid or pseudotyping env plasmid) comprising a gene encoding an envelope protein(s) (such as

VSV-G), c) a plasmid vector comprising between 5' and 3' LTR sequences, a psi encapsidation sequence, and a transgene, and d) a plasmid vector comprising a nucleic acid construct encoding the first and second proteins (either separately or as part of the fusion protein described above) disclosed herein. In some embodiments, the nucleic acid construct encoding the first and second proteins (either separately or as part of the fusion protein described above)disclosed herein is on the packaging plasmid instead of a separate plasmid. Nucleic acids encoding gag, pol and env cDNA can be advantageously prepared according to conventional techniques, from viral gene sequences available in the prior art and databases.

In some embodiments, a lentiviral vector comprises a nucleic acid construct as described herein. In some embodiments, a lentiviral vector comprises the first and second proteins (either separately or as part of the fusion protein described above) as described herein.

The promoters used in the plasmids can be identical or different. In some embodiments, in the plasmid transcomplementation system, the envelope plasmid and the plasmid vector, respectively, to promote the expression of gag and pol of the coat protein, the mRNA of the vector genome and the transgene are promoters which can be identical or different. Such promoters can be chosen advantageously from ubiquitous promoters or specific, for example, from viral promoters CMV, TK, RSV LTR promoter and the RNA polymerase III promoter such as U6 or H1 or promoters of helper viruses encoding env, gag and pol (i.e. adenoviral, baculoviral, herpes viruses).

For the production of the lentiviral vector of the disclosure, the plasmids described herein can be introduced into host cells and the viruses are produced and harvested. Suitable cells include but not limited to eukaryotic and prokaryotic cells and/or cell lines. Non-limiting examples of such cells or cell lines generated from such cells include, e.g., COS, CHO (e.g., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (e.g., HEK293-F, HEK293-H, HEK293-T), and perC6 cells as well as insect cells such as Spodoptera fugiperda (Sf), or fungal cells such as Saccharomyces, Pichia and Schizosaccharomyces.

Once host cells are transfected with the plasmids and a lentiviral vector (or particles) of the disclosure is produced, the lentiviral vectors (or particles) of the disclosure can be purified from the supernatant of the cells. Purification of the lentiviral vector to enhance the concentration can be accomplished by any suitable method, such as by density gradient purification (e.g., cesium chloride (CsCl)), by chromatography techniques (e.g., column or batch chromatography), or by ultracentrifugation. For example, the vector of the invention can be subjected to two or three CsCl density gradient purification steps. The vector, is desirably purified from infected cells using a method that comprises lysing cells, applying the lysate to a chromatography resin, eluting the virus from the chromatography resin, and collecting a fraction containing the lentiviral vector of the disclosure.

Methods of delivery of lentiviral vectors have been described. See, e.g., Vargas et al. J. Trans. Med. 14(288):1-15 (2016); Mali et al. Nat. Methods 10(10):957-963; Mátés et al. Nature Genetics 41(6):753-761 (2009); Skipper et al. 20(92):1-23 (2013).

Lentiviral vectors comprising the first and second proteins (either separately or as part of the fusion protein described above) or a nucleic acid construct coding therefor can be administered to a subject by any route. In some embodiments, a lentiviral vector of the disclosure can be delivered to cells of a subject either in vivo or ex vivo.

In some embodiments, the lentiviral vector of the disclosure can be delivered in vivo. In some embodiments, a lentiviral vector comprising the first and second proteins (either separately or as part of the fusion protein described above) encoded by a nucleic acid construct of the disclosure can be used to deliver a gene of interest and/or to target a genetic defect in a subject's DNA. In some embodiments, the lentiviral vector is administered to the subject parenterally, preferably intravascularly (including intravenously). When administered parenterally, it is preferred that the vectors be given in a pharmaceutical vehicle suitable for injection such as a sterile aqueous solution or dispersion.

In some embodiments, the lentiviral vector of the disclosure can be used ex vivo.

In some embodiments, a lentiviral vector comprising the first and second proteins (either separately or as part of the fusion protein described above) encoded by a nucleic acid construct of the disclosure can be used to deliver a gene of interest and/or target a genetic defect in a subject's DNA. In some embodiments, cells are removed from a subject and lentiviral vector comprising the first and second proteins (either separately or as part of the fusion protein described above) encoded by a nucleic acid construct of the disclosure is administered to the cells ex vivo to modify the DNA of the cells. The cells carrying the modified DNA are then expanded and reinfused back into the subject. In certain embodiments, a lentiviral vector comprising the first and second proteins (either separately or as part of the fusion protein described above) encoded by a nucleic acid construct of the disclosure can be used for Chimeric Antigen Receptor (CAR) T-cell therapy to genetically modify a patient's autologous T-cells to express a CAR specific for a tumor antigen. In a further embodiment, the modified CAR-T cells are expanded ex vivo and re-infusion back to the patient. In some embodiments, the altered T cells more specifically target cancer cells. Unlike antibody therapies, CAR-T cells are able to replicate in vivo resulting in long-term persistence.

Following administration of a lentiviral vector of the disclosure or cells modified ex vivo using a lentiviral vector of the disclosure, the subject can be monitored to detect the expression of the transgene. Dose and duration of treatment is determined individually depending on the condition or disease to be treated. A variety of conditions or diseases can be treated based on the gene expression produced by administration of the gene of interest in the vector of the present invention. The dosage of vector delivered using the method of the invention will vary depending on the desired response by the host and the vector used.

In some gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Accordingly, a viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest.

Certain aspects of the disclosure are directed to a method of inserting an exogenous nucleic acid sequence into genomic DNA of an organism, comprising: identifying the specific genomic DNA sequence in the genome of the organism; administering a lentiviral particle comprising the nucleic acid construct of the disclosure to the organism to bind to the specific genomic DNA sequence and insert the exogenous nucleic acid into the genomic DNA; wherein the exogenous nucleic acid becomes integrated at the specific genomic DNA sequence.

Certain aspects of the disclosure are directed to a method for controlled, site-specific integration of a single copy or multiple copies of an exogenous nucleic acid sequence into a cell, the method comprising: a) delivering the nucleic acid construct, the vector, or the first and second proteins (either separately or as part of the fusion protein described above) of the disclosure to the cell, and b) delivering the exogenous nucleic acid to the cell; wherein binding of the first and second proteins (either separately or as part of the fusion protein described above) to the specific genomic DNA sequence in the genome of the cell, results in cleavage of the genome and integration of one or more copies of the exogenous nucleic acid into the genome of the cell. In some aspects, the delivery to the cell is by means of a lentiviral particle.

Several strategies can be used to test for integrations sites, and to screen for the best machinery for directed integration.

For analysis of the modified transposons disclosed herein, a reporter cell line with a promoter, half of the coding sequence of the GFP and a splice site donor downstream of the targeted insertion site in the genome can be used. For example, the lentiviral payload can have a fusion integrase variant followed by the inverted splice site acceptor and the other half of the GPF. The expression of GFP will occur when direct insertion happens and splicing of the GFP containing mRNA generated from the insertion site and integrated payload originates the full GFP CDS.

VPR transcomplementation systems can also be used for screening and comparing integration mutants. The transcomplementation system can be used for targeted insertion

of the lentiviral payload containing a fusion integrase variant that, when expressed and loaded in the particle promote its own integration will be loaded in the viral particle using a VPR fusion. This will complement in trans the integration defective IN coded in the packaging vector used for particle production. Other methods that can be used for integration mapping including IC, or FISH probes. Targeted insertion can also be screened by TCRa or RFP targeted disruption, or GFP activation by targeted splice site integration.

For the FISH approach to co-staining of the insertion and target region in the chromatin, a Fluorescence in situ hybridization to localize the gene of interest transposon in the Hek293T genome can be performed. Hek293T can be transfected with 1) GOI-transposon 2) Programmable transposase and 3) gRNA to PPP1R12. Probes are designed to target the PPP1R12 gene, CD46 gene (as negative control) and GOI, and can be synthesized with Nick Translation Mix (Sigma) from PCR amplified DNA.

In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) comprising a modified transposase as disclosed herein improve the specificity of insertion of the exogenous nucleic acid into the genome compared to a wild-type transposase (or a fusion protein containing the corresponding wildtype transposase), e.g., as determined by a Genetrap assay. In some embodiments, HEK293T cells, or any other permissible cells, are transfected or transduced with lentiviral particles with the following plasmids or payloads: (i) a plasmid comprising a gRNA that targets a specific region of DNA, (ii) a plasmid comprising the nucleic acid construct of the disclosure encoding the first and second proteins (either separately or as part of the fusion protein described above) with the second protein being a modified transposase, and (iii) a genetrap plasmid comprising a nucleic acid sequence encoding a reporter protein, e.g., GFP, that lacks a promoter. In some embodiments, the genetrap plasmid further comprises a transposon with inverted repeats.

In some embodiments, the percent of cells containing the GFP insertion can be determined by flow cytometry. In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) with the second protein being a modified transposase increase the percent of cells containing insertion of GFP by at

least 5%, at least 10%, at least 15%, at least 20%, at least 25%, or at least 30% compared to the corresponding wildtype protein. In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) with the second protein being a modified transposase increase the percent of cells containing insertion of GFP by about 15-30%.

In some embodiments, the percent of insertions at the targeted site and percent of coverage at the target site (number of reads per insertion site) can be determined by genomic DNA extraction and targeted sequencing with oligonucleotides specific for viral LTRs. In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) with the second protein being a modified transposase increase the percent of insertions at the targeted site by at least 10-fold, at least 20-fold, at least 30-fold, at least 40-fold, at least 50-fold, at least 60-fold, at least 70-fold, at least 80-fold, at least 90-fold, or at least 100-fold compared to the corresponding wildtype protein. In some embodiments, the percent of insertions at the targeted site is increased by about 10-100 fold. In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above) with the second protein being a modified transposase increase the percent of coverage at the target site (number of reads per insertion site) by at least 10-fold, at least 20-fold, at least 30-fold, at least 40-fold, at least 50-fold, at least 60-fold, at least 70-fold, at least 80-fold, at least 90-fold, at least 100-fold, at least 110-fold, at least 120-fold, at least 130-fold, at least 140-fold, at least 150-fold, at least 160-fold, at least 170-fold, at least 180-fold, at least 190-fold, or at least 200-fold compared to the corresponding wildtype protein. In some embodiments, the percent of coverage at the target site (number of reads per insertion site) by at least 100-fold.

In some embodiments, the percent of insertions at the targeted site and percent of coverage at the target site (number of reads per insertion site) can be determined by genomic DNA extraction and targeted sequencing with oligonucleotides specific for viral inserted LTR.

Possible applications of lentiviral vectors comprising the first and second proteins (either separately or as part of the fusion protein described above) of the disclosure include gene

therapy, i.e., the gene transfer in any mammal cell, in particular in human cells. It may be dividing cells or quiescent cells, cells belonging to the central organs or peripheral organs such as the liver, pancreas, muscle, heart, etc. Gene therapy may allow the expression of proteins, e.g. neurotrophic factors, enzymes, transcription factors, receptors, etc. Lentiviral vectors according to the invention may also particularly suitable for research purposes.

In some embodiments, a nucleic acid constructs, the first and second proteins (either separately or as part of the fusion protein described above), and/or a lentiviral vector of the disclosure is administered to a subject to treat a disease. In some embodiments, the disease is a genetic disorder that can benefit from gene therapy.

In some embodiments, the first and second proteins (either separately or as part of the fusion protein described above), or the nucleic acid constructs coding therefor, or the kit or composition as disclosed hereafter or the lentiviral vectors comprising the first and second proteins (either separately or as part of the fusion protein described above) or nucleic acid constructs according to the disclosure can be used as a medicament.

The lentiviral vector according to the disclosure may be particularly suitable for treating a genetic disease in a subject.

The present invention also relates to a composition comprising

(i) a RNA guided nuclease or zinc finger nuclease,

(ii) a transposase,

(iii) a guide RNA, and

(iv) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome,

wherein said transposase is a modified hyperactive Piggybac, comprising one or more amino acid mutations as compared to hyperactive Piggybac of SEQ ID NO:9.

In one embodiment, the modified hyperactive PiggyBac mutation comprises the amino acid substitution R372A/K375A/D450N.

In a preferred embodiment, the modified hyperactive PiggyBac mutation does not comprise the amino acid substitution R372A/K375A/D450N.

In some embodiments, the modified hyperactive PiggyBac mutation comprises the following amino acid substitution or combination of amino acid substitution of S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, T43I/R372A/K375A/A411T/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In a preferred embodiment, the modified hyperactive PiggyBac mutation comprises the following amino acid substitution or combination of amino acid substitution of R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N,

G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

In some embodiments the RNA guided nuclease is a Cas9 protein. In some embodiments the RNA guided nuclease is a SpCas9 protein. In some embodiments the RNA guided nuclease is a SaCas9 protein.

The present invention also relates to a composition comprising nucleic acids encoding:

(i) a RNA guided nuclease or zinc finger nuclease, as described hereinabove,

(ii) a transposase, as described hereinabove,

(iii) a guide RNA, and

(iv) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome.

In some embodiments, the nucleic acids of the composition are expressed in a cell through a suitable expression vector. As used herein, the term "expression vector" refers to a vector comprising a polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an in vitro expression system. Expression vectors include all those known in the art, including cosmids, plasmids (e.g., naked or contained in liposomes) and viruses (e.g., lentiviruses, retroviruses, adenoviruses, and adeno-associated viruses) that incorporate the recombinant polynucleotide.

In a preferred embodiment, the two nucleic acids are co-expressed in the same cell or cell population. In some embodiments, the two nucleic acids are co-expressed concomitantly. In another embodiment, the nucleic acid encoding an RNA guided nuclease or zinc finger nuclease is expressed first. In another embodiment, the nucleic acid encoding a transposase is expressed first.

The invention further relates to a composition comprising

(i) a fusion protein comprising a RNA-guided nuclease and a transposase as disclosed herein, or a nucleic acid encoding thereof,

(ii) a transposase as disclosed herein, or a nucleic acid encoding thereof,

(iii) a guide RNA, and

(iv) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome.

The invention further relates to a composition comprising

(i) a first fusion protein comprising a RNA-guided nuclease and a transposase as disclosed herein, or a nucleic acid encoding thereof,

(ii) a second fusion protein comprising a RNA binding protein engineered to bind to at least one specific RNA sequence, and a transposase as disclosed herein, or a nucleic acid encoding thereof,

(iii) a guide RNA, and

(iv) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome.

The present disclosure also provides compositions for practicing the disclosed methods as described herein. In some embodiments, a composition comprises a nucleic acid construct or a vector as defined in this disclosure, and a polynucleotide sequence encoding an exogenous nucleic acid for insertion in a genome, contained in in or bound to a packaging vector.

The present disclosure further relates to a composition comprising

(i) a fusion protein comprising a RNA-guided nuclease and a transposase as disclosed herein or a nucleic acid encoding said fusion protein,

(ii) a guide RNA, and

(iii) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome.

In specific embodiments, said nucleic acid or gene of interest is a large DNA fragment, typically having a size between 5kb and 25kb, and more preferably between 8kb and 20kb.

Also provided by the present disclosure are kits for practicing the disclosed methods, as described herein. The kit can contain the nucleic acid constructs or fusion proteins as described herein. In some aspects, the kit can contain the lentiviral particles containing the nucleic acid constructs or fusion proteins as described herein.

The subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions can be printed on a substrate, such as paper or plastic, etc. As such, the instructions can be present in the kit as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging), etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g., via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

The disclosure typically relates to a kit, comprising

a first composition including

(i) a first fusion protein as defined herein, or a nucleic acid encoding said first fusion protein, and wherein said first fusion protein comprises an amino acid sequence of a first guided RNA nickase Cas9, typically SpCas9 nickase of SEQ ID NO:70 fused to a modified hyperactive Piggybac, and,

(ii) a first guided RNA nucleic acid,

a second composition including

(iii)a second fusion protein as defined herein, or a nucleic acid encoding said second fusion protein, and wherein said second fusion protein comprises an amino sequence of a second guided RNA nickase Cas9, typically SaCas9 nickase of SEQ ID NO:76 fused to a modified hyperactive Piggybac,

(iv)a second guided RNA nucleic acid,

optionally, an nucleic acid for insertion in a genome, for example a nucleic acid having a size between 5kb and 25kb, and more preferably between 8kb and 20kb.

wherein said first and second fusion protein are capable of forming heterodimerization and making double cuts determined by said first and second guided RNA at adjacent sites of a genomic DNA region, and optionally inserting said nucleic acid between the adjacent sites.

In specific embodiments, the composition or kit comprises exogenous nucleic acid in a minicircle, a plasmid or a viral vector, in particular in non-integrating viral vector, for example or non-integrating lentiviral vector.

In specific embodiments, the composition or kit as disclosed herein is comprised in a nanoparticle.

In specific embodiments, said composition is a nucleic acid composition comprising

(i) a nucleic acid construct encoding the fusion protein as disclosed herein,

(ii) a guide RNA, and

(iii) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome.

In specific embodiments, said kit comprising

- a first composition including

(i) a nucleic acid construct encoding a first fusion protein as disclosed herein, and wherein said first fusion protein comprises an amino acid sequence of a first guided RNA nickase Cas9, typically SpCas9 nickase of SEQ ID NO:70 fused to a modified hyperactive Piggybac, and,

(ii) a first guided RNA nucleic acid,

- a second composition including

(iii) a nucleic acid construct encoding said second fusion protein as disclosed herein, and wherein said second fusion protein comprises an amino sequence of a second guided RNA nickase Cas9, typically SaCas9 nickase of SEQ ID NO:76 fused to a modified hyperactive Piggybac, and,

(iv) a second guided RNA nucleic acid,

In specific embodiments, said kit or composition is for use as a drug, in particular in treating disorders in human, for example for treating genetic deficiencies in a human subject in need thereof.

In some embodiments, the nucleic acid construct is in form of RNA, DNA or protein, and the polynucleotide sequence encoding the exogenous nucleic acid is in form of RNA or DNA, depending on the method of delivery. Particularly, the polynucleotide sequence encoding the exogenous nucleic acid is in form of RNA.

In some embodiments, the composition or kit is viral-free and the packaging vector is a nanoparticle e.g. a polymeric or lipidic nanoparticle. The packaging vector can also be a carrier which is bound to the elements of the composition. In some embodiments, the composition is contained in a viral vector, particularly a lentiviral particle.

In some embodiments, the composition or kit comprises (a) the nucleic acid construct described herein (e.g. comprising Cas9 and a transposase) in form of RNA, (b) a guide RNA if needed (e.g. as separate lineal single strand RNA molecule), and (c) a polynucleotide comprising the exogenous gene for insertion in DNA form (e.g. in a vector), contained in in or bound to a packaging vector.

In some embodiments, the composition comprises (a) the fusion protein described herein (e.g. comprising Cas9 and a transposase) in form of protein, (b) a guide RNA if needed (e.g. as separate lineal single strand RNA molecule), wherein the fusion protein and the guide RNA form a ribonucleic protein complex (RNP), and (c) a polynucleotide comprising the exogenous gene for insertion in DNA form (e.g. in a vector), contained in in or bound to a packaging vector.

In some embodiments, the composition comprises (a) the nucleic acid construct described herein (e.g. comprising Cas9 and a transposase) in form of DNA, (b) a guide RNA if needed (e.g. as separate lineal RNA molecule or as DNA in a vector), and (c) a polynucleotide comprising the exogenous gene for insertion in DNA form (e.g. in a vector), contained in in or bound to a packaging vector.

In some embodiments, the composition comprises (a) the fusion protein described herein (e.g. comprising Cas9 and an integrase) in form of protein, (b) a guide RNA if needed (e.g. as separate RNA molecule complexing with the fusion protein), and (c) a polynucleotide comprising the exogenous gene for insertion, contained in in or bound to a packaging vector. In a particular embodiment, the packaging vector is a lentiviral particle. In some embodiments, the (a) fusion protein is bound to the lentiviral capside by means of gag-pol or VPR (Viral Protein R). In some embodiments, the (c) polynucleotide is in form of RNA as payload of the integrase.

In a particular embodiment, when ZFP is used, (b) the guide RNA can not be needed.

The invention further relates to a composition comprising

(i) a fusion protein comprising a RNA binding protein engineered to bind to at least one specific RNA sequence, a DNA binding protein enabling the insertion of an exogenous nucleic acid into the genome, and a linker connecting the first and second protein,

(ii) a Cas9 protein, and

(iii) a guide RNA comprising said at least one specific RNA sequence for fusion protein binding,

(iv) a nucleic acid or gene of interest, for example an exogenous nucleic acid for insertion in a genome

wherein the DNA binding protein is a modified transposase of this disclosure, typically a modified hyperactive Piggybac comprising one or more amino acid mutations to increase excision activity as compared to unmodified hyperactive Piggybac, and one or more amino acid mutations to decrease DNA binding activity as compared to unmodified hyperactive Piggybac according to SEQ ID NO: 9,

In some embodiments, the RNA binding protein is the MS2 bacteriophage coat protein (MCP).

In some embodiments, the at least one RNA sequence recognized by the MCP of the fusion protein is a tetraloop. As used herein, the term “tetraloop” is used interchangeably

with the terms “stem loop” and “hairpin loop”. In some embodiments, the at least one RNA tetraloop is a MS2 RNA tetraloop binding sequence.

In some embodiments the guide RNA comprises at least one MS2 RNA tetraloop binding sequence. In some embodiments, the gRNA comprises more than one MS2 RNA tetraloop binding sequences. As used herein, the term “more than one” means 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or more.

Various embodiments are described in the claims. Some additional embodiments are disclosed hereafter:

E1. A fusion protein comprising

(i) a first protein consisting of an RNA guided nuclease or zinc finger nuclease,

(ii) a second protein consisting of a transposase, and

(iii) optionally, a linker connecting the first and second protein,

wherein said transposase is a modified hyperactive Piggybac, comprising one or more amino acid mutations as compared to hyperactive Piggybac of SEQ ID NO:9,

and wherein said first protein is fused at the C-terminal end of the second protein directly, or indirectly via a linker.

E2. The fusion protein of Embodiment 1, wherein said transposase is a modified hyperactive Piggybac, comprising one or more amino acid mutations to increase excision activity as compared to unmodified hyperactive Piggybac, and one or more amino acid mutations to decrease DNA binding activity as compared to unmodified hyperactive Piggybac,

E3. The fusion protein of Embodiment 2, wherein said one or more amino acid mutations to increase excision activity are selected among the amino acid mutations within the region defined by the amino acid position numbers [194-200], [214-222], [434-442] or [446-456], for example amino acid substitution at the position D198, D201, R202, M212, or S213; said position number

corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9.

E4. The fusion protein of any one of Embodiments 1-3, wherein said one or more amino acid mutations are selected among the amino acid substitutions which increase excision activity at position of M194 or D450, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9, preferably selected among the amino acid substitutions M194V and/or D450N.

E5. The fusion protein of any one of Embodiments 1 to 4, wherein said one or more amino acid mutations are selected among the amino acid substitutions which decrease DNA binding activity at position R372, K375, R376, E377, and/or E380, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9, preferably selected among among the amino acid substitutions R372A, K375A, R376A, E377A, and/or E380A.

E6. The fusion protein of any one of Embodiments 1 to 5, wherein the modified hyperactive Piggybac includes at least one amino acid substitution to increase excision activity at position D450, and at least two amino acid substitutions to decrease DNA binding activity at position R372 and K375, preferably said modified transposase of hyperactive Piggybac includes the triple mutations D450N, R372A and K375A, said position number corresponding to the amino acid number of unmodified hyperactive Piggybac of SEQ ID NO:9.

E7. The fusion protein of any one of Embodiments 1 to 6, wherein the modified hyperactive Piggybac further comprises at least one mutation in the region defined by the amino acid position numbers [158-169], for example A166S; and/or at least one mutation at position Y527, R518, K525, N463.

E8. The fusion protein of any one of Embodiments 1 to 7, wherein said modified hyperactive Piggybac comprises an amino acid sequence having at least 85%, at least 90%, at least 95% identity, or 100% identity to the modified hyperactive Piggybac of SEQ ID NO:1.

E9. The fusion protein of any one of Embodiments 1 to 8, wherein said modified hyperactive Piggybac is a variant of the hyperactive Piggybac of SEQ ID NO:1 with one or more amino acid substitutions, typically with no more than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions.

E10. The fusion protein of any one of Embodiments 1 to 9, wherein said modified hyperactive Piggybac further comprises one or more amino acid substitutions at the position number 245, 268, 275, 277, 287, 290, 315, 325, 341, 346, 347, 350, 351, 356, 357, 388, 409, 412, 432, 447, 460, 461, 465, 517, 560, 564, 571, 573, 576, 586, 587, 589, 592, and/or 594, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

E11. The fusion protein of Embodiment 10, wherein the modified hyperactive PiggyBac mutation comprises the following amino acid substitution or combination of amino acid substitution of R245A, D268N, R275A, R277A, K287A, K290A, K287A/K290A, R315A, G325A, R341A, D346N, N347A, N347S, T350A, S351E, S351P, S351A, K356E, N357A, R388A, K409A, K412A, K432A, D447A, D447N, D450N, R460A, K461A, W465A, S517A, T560A, S564P, S571N, S573A, K576A, H586A, I587A, M589V, S592G, or F594L, D450N/R372A/K375A, R275A/R277A, K409A/K412A, R460A/K461A, R275A/R277A/N347S/K375A/T560A/S573A/M589V/S592G and R245A/R275A/R277A/R372A/W465A, the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9).

E12. The fusion protein of Embodiment 10, wherein the modified hyperactive PiggyBac mutation comprises the following amino acid substitution or combination of amino acid substitution of R372A/K375A/D450N, R372A/K375A/R376A/D450N, K375A/R376A/E377A/E380A/D450N, R372A/K375A/R376A/E377A/E380A/D450N, M194V, R376A, E377A, E380A M194V/R372A/K375A, S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L,

R202K/R275A/N347S/R372A/D450N/T560A/F594L,R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L ; the position number corresponding to the amino acid number of the hyperactive PiggyBac sequence (SEQ ID NO: 9) typically said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 1-8 and 10-18.

E13. The fusion protein of any one of Embodiments 1-12, wherein the linker is a peptidic linker which comprises a XTEN sequence or a GGS sequence, preferably a XTEN sequence.

E14. The fusion protein of any one of Embodiments 1-13, wherein the linker is a peptidic linker having between 3 to 50 amino acids in length, typically selected among any of SEQ ID NO: 49, 51, 53, 55, 57, 59, 61.

E15. The fusion protein of any one of Embodiments 1-14, wherein said first protein is a Cas9 protein comprising an active DNA cleavage domain and a guide RNA binding domain.

E16. The fusion protein of any one of Embodiments 1-15, wherein said first protein is a nuclease protein comprising an active DNA cleavage domain and a guide RNA binding domain and having at least 80%, 90%, 95%, 99% or at least 100% identity to a Streptococcus pyogenes Cas9 of SEQ ID NO:31, SaCas9 of SEQ ID NO:72, Cpf1 of SEQ ID NO:74, CjCas9 of SEQ ID NO:29, SpCas9 nickase of SEQ ID NO:70, CasX of SEQ ID NO:75, or SaCas9 nickase of SEQ ID NO:76.

E17. The fusion protein of any one of Embodiments 1-16, wherein said first protein is a Cas9 protein selected from the group consisting of a SaCas9 of SEQ ID NO:72 or Streptococcus pyogenes Cas9 of SEQ ID NO:31.

E18. The fusion protein of any one of Embodiments 1 to 17, which is a triple fusion protein comprising

(i) a first protein consisting of an RNA guided nuclease or nickase,

(ii) a second protein consisting of a first transposase,

(iii)a third protein consisting of a second transposase and,

(iv)optionally, peptidic linkers between the first and second protein and the second and the third protein,

and wherein the first and second transposases have identical or different sequences of modified Piggybac transposase, for example as defined in any one of Embodiments 3-12.

E19. The fusion protein of any one of Embodiments 1 to 14, wherein said first protein is a zinc finger protein of SEQ ID NO:33.

E20. A composition comprising

(i) a fusion protein of any of Embodiments 1-19, or a nucleic acid encoding said fusion protein,

(ii) a guide RNA, and

(iii) an exogenous nucleic acid for insertion in a genome.

E21. The composition of Embodiment 20, wherein said exogenous nucleic acid is a large DNA fragment, typically having a size between 5kb and 25kb, and more preferably between 8kb and 20kb.

E22. A kit, comprising

(i) a first composition including

- a first fusion protein as defined in any one of Embodiments 1 and 18 or a nucleic acid encoding said first fusion protein, and wherein said first fusion protein comprises an amino acid sequence of a first guided RNA nickase Cas9, typically SpCas9 nickase of SEQ ID NO:70 fused to a modified hyperactive Piggybac,

- a first guided RNA nucleic acid,

(ii) a second composition including

- a second fusion protein as defined in any one of Embodiments 1 to 18, or a nucleic acid encoding said second fusion protein, and wherein said second fusion protein comprises an amino sequence of a second guided RNA nickase

Cas9, typically SaCas9 nickase of SEQ ID NO:76 fused to a modified hyperactive Piggybac,

- a second guided RNA nucleic acid,

(iii) optionally, an exogenous nucleic acid for insertion in a genome, for example an exogenous nucleic acid having a size between 5kb and 25kb, and more preferably between 8kb and 20kb.

wherein said first and second fusion proteins are capable of forming heterodimerization and making double cuts determined by said first and second guided RNA at adjacent sites of a genomic DNA region, and optionally inserting said exogenous nucleic acid between the adjacent sites.

E23. The composition or kit of Embodiments 20-22, wherein said exogenous nucleic acid is comprised in a minicircle, a plasmid or a viral vector, in particular non-integrating viral vector, for example or non-integrating lentiviral vector.

E24. The composition or kit of Embodiments 20-23, wherein the compositions are comprised in a nanoparticle.

E25. A modified hyperactive Piggybac transposase, comprising at least one amino acid mutation to increase excision activity as compared to unmodified hyperactive Piggybac, and/or at least one amino acid mutation to decrease DNA binding activity as compared to unmodified hyperactive Piggybac,

wherein said at least one mutation to increase excision activity is an amino acid substitution of M at position 194, typically M194V and/or wherein at least one amino acid mutation to decrease DNA binding activity are selected among the amino acid substitutions at positions R376, E377, and E380, typically R376A, E377A, and/or E380A.

E26. A modified hyperactive Piggybac transposase, which comprises the following combination of mutations R372A/K375A/D450N, R372A/K375A/R376A/D450N, K375A/R376A/E377A/E380A/D450N, R372A/K375A/R376A/E377A/E380A/D450N, M194V, R376A, E377A, E380A, M194V/R372A/K375A,

S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L,R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L as compared to unmodified hyperactive Piggybac of SEQ ID NO:9.

E27. The modified hyperactive Piggybac transposase of Embodiment 25 or 26, further comprising one mor more amino acid mutations selected from the group consisting: amino acid at position 245 is A, amino acid at position 275 is R or A, amino acid at position 277 is R or A, amino acid at position 325 is A or G, amino acid at position 347 is N or A, amino acid at position 351 is E, P or A, amino acid at position 372 is R, amino acid at position 375 is A, amino acid at position 450 is D or N, amino acid at position 465 is W or A, amino acid at position 560 is T or A, amino acid at position 564 is P or S, amino acid at position 573 is S or A, amino acid at position 592 is G or S, and amino acid at position 594 is L or F.

E28. The modified hyperactive PiggyBac transposase of Embodiment 25 or 26, which comprises 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 additional mutations as compared to unmodified hyperactive PiggyBac of SEQ ID NO: 9, wherein the modified hyperactive PiggyBac shows decreased DNA binding activity and/or increased excision activity compared to hyperactive PiggyBac of SEQ ID NO:9.

E29. A nucleic acid encoding the fusion protein of any one of Embodiments 1 to 18, typically a messenger RNA (mRNA).

E30. The nucleic acid encoding the fusion protein of Embodiment 28, which comprises a sequence selected from the group consisting of SEQ ID NO: 110-112 or their corresponding mRNA sequence.

E31. A nucleic acid encoding the modified hyperactive Piggybac of any one of Embodiments 25-28.

E32. An expression vector comprising the nucleic acid of any one of Embodiments 29-31.

E33. A host cell comprising the nucleic acid of any one of Embodiments 29-31, or the expression vector of Embodiment 32.

E34. A method for site specific integration of an exogenous nucleic acid sequence into the genome of a cell, the method comprising delivering to the cell, a composition comprising

(i) a fusion protein of any one of Embodiments 1 to 18, or a nucleic acid of any one of Embodiments 29-31,

(ii) an exogenous nucleic acid to be integrated into the genome of the cell, and

(iii) a guide RNA for determining site-specific integration of said exogenous nucleic acid into the genome of the cell.

wherein binding of said fusion protein to the specific genomic DNA sequence in the genome of the cell results in cleavage of the genome and site-specific integration of said exogenous nucleic acid sequence into the genome of the cell as determined by the guide RNA.

E35. The method of Embodiment 34, wherein said exogenous nucleic acid is a nucleic acid fragment of a size of at least 5kb, at least 6kb, at least 7kb, at least 8 kb, at least 9kb, typically comprised between 5 and 25 kb, preferably between 8 and 20 kb.

E36. The method of Embodiment 34 or 35, wherein said exogenous nucleic acid is a therapeutic transgene to be inserted in a genome of a subject in need thereof to correct the deficiency of a genetic disorder.

E37. The method of any one of Embodiments 34-36, wherein said composition is delivered in vitro or ex vivo, typically in a mammalian cell, preferably a human cell, and more preferably in a human cell which have been obtained from a human subject suffering from a genetic disorder.

E38. The method of any one of Embodiment 34-37, wherein said composition is delivered in vivo into a mammal, for example a human subject in need thereof, typically for therapeutic treatment of a genetic disorder.

E39. A method of inserting an exogenous nucleic acid sequence into genomic DNA of an organism, comprising: administering one or more compositions or kits as defined in any one of Embodiment 20-24, to the organism such that the fusion protein comprised in said one or more compositions or kits bind to a specific genomic DNA sequence and enables the insertion of the exogenous nucleic acid comprised in said composition into the genomic DNA;

wherein the exogenous nucleic acid becomes integrated at a specific site of the genome of a cell of said organism, for example, a non-human organism or a human subject in need thereof.

EXAMPLES

To achieve a programmable transposase, our principle of design attempted to combine the precise genome-targeting of CRISPR systems with PB variants that exhibit enhanced insertion and excision activity and lower target DNA binding activity. We started by fusing a nuclease SpCas9 with a PB transposase (Fig. 1). We constructed a diverse library of PB and SpCas9 variants (Fig. 2a, b, c) where we tested a library of PB variants, 3 SpCas9 variants (nuclease SpCas9 (cas9), dead SpCas9 (dcas9), nickase SpCas9 (ncas9)), and 6 linkers (4xGGS, 5xGGS, 7xGGS, 8xGGS, XTEN, FOKI).

Different parts of the PB transposase were diversified. Special emphasis focused on the catalytic core domain which is formed by a catalytic triad of 3 aspartate residues (D268, D346, D447) surrounded by 15 arginine and lysine residues likely involved in transposase-target DNA interactions[14]. In addition to the hyperactive versions explored in the past[9] we diversified additional residues that may influence PB excision and integration activities[15] (SEQ 1-N). In order to isolate the best performing mutants, we developed a sensitive reporter system for targeted gene insertion, based on Emerald GFP (emGFP) reconstitution upon on-target integration. A promoterless C-terminal (C-t) half

of emGFP preceded by a splicing acceptor was inserted in the genome of Hek293T cells to build a reporter cell line (dubbed as Hershey). A complementary 'insertion-trap reporter' was constructed encoding the N-terminal (N-t) first half of a emGFP with an upstream promoter followed by a splice donor and all flanked by the PB inverted repeats. Specific gRNA-guided cas9 directed PB to insert the N-t half adjacent to the C-t half, which upon splicing of the resulting transcript leads to production of green fluorescence (Fig. 3). A library of cas9-PB chimeric proteins were assembled and transfected with the 'insertion-trap reporter' and guide-RNA (gRNA) to the cells containing the Hershey reporter. The variants presenting higher programmable insertion were tested separately using the same reporter cell line (Figure 2a, 2b; 4a, 4b). This assay tested on-target activity (emGFP positive cells) and total transposition activity (RFP positive cells) using a dual transposon containing a N-t half emGFP and a full RFP sequence upstream. On-target to off-target ratio was calculated by dividing the percentage of emGFP positive cells (on-target activity) to the total percentage of RFP positive cells (total insertion activity).

We tested various combinations of cas9 variants fused to PB variants for on-target and off-target transposition in the Hershey cell line. We observed the highest levels of programmable insertion in Nter-cas9-PB-Cter fusions containing i) cas9 with intact nuclease activity (Fig. 2A) and ii) PB variants with increased excision activity and iii) decreased t-DNA binding activity (Fig. 2B; 4A, 4B). We observed that all three parameters are important for achieving precise and efficient targeted insertion. First, the requirement for cas9 nuclease activity was demonstrated by the significantly lower efficiency of targeted insertion by ncas9 (D10A) and dcas9 (D10A and H840A) fused to PB (Fig. 2A). To further explore the role of the double-strand break (DSB) activity of cas9 in facilitating targeted integration, we uncoupled on-site targeting and DSB activity by using a Zinc finger-PB fusion for directed localization of the transposon and complemented it with on-site DSB by an independent Cas9 nuclease. Znf-PB fusion exhibited no or very low targeted insertion activity that was rescued when combined with introducing DSBs near the Znf binding site with gRNA guided-cas9 (Fig. 5). These results are consistent with a mechanism where DSB generation by cas9 in the vicinity of PB facilitates the insertional activity of PB and bypasses its requirement for the canonical

TTAA at the insertion site. Indeed, our analysis of the cas9-PB insertion sites showed that Inverted terminal repeat (ITR) sequences get disrupted with the presence of small indels near the targeting site (Fig. 6). An important consequence of this disruption is the irreversibility of the PB-mediated integration mechanism. Mobilization of transposons with disrupted ITRs or TTAA is either eliminated or reduced[16]. This mechanism likely contributes to the efficiency of programmable insertion by cas9-PB, and the coupling of 'find' and 'cut' activity of cas9 with 'pasting' activity of modified PB contributes high levels of precision. Second, high excision PB mutants appear to contribute to higher programmable insertion (D450N, M194V; Fig. 2B, 4A, 4B). This enhanced excision may be a result of increased donor DNA substrate primed for integration. Furthermore, the destruction of the preferred substrate of excision may prevent enough excision even from PB mutants with higher excision activity (Fig. 6). Third, PB mutants with reduced t-DNA binding resulted in the lowest off-target levels (Fig. 2B, 4A, 4B). This result is consistent with a decrease in the intrinsic non-specific DNA binding by PB that is complemented by the sequence-specific DNA binding of cas9 in the cas9-PB fusion which inactivates PB solo insertion activity (Figure 7).

Using a genome guide approach, we were able to characterize programmable transposase insertion sites and off-target levels using a modified version of a Guide-seq[17] based protocol (Fig. 8A). None of the on-target insertions detected happened at TTAA sites, further demonstrating integration on DSB sites generated by cas9 and resulting in the loss of preferred excision substrate (Fig. 8B). Additionally we run Guide-seq analysis on cells modified with programmable transposase technology targeting the TRAC loci. We detected all insertions on-target with sensitivity down to 1-10% (Fig. 9).

We have benchmarked programmable transposase technology with current methods for precise gene delivery such as cas9 based HDR (Fig. 10) programmable transposase shows higher efficiencies, a gap which widens in large payloads. The best mutants achieved insertions (up to 8kb) with 2-fold more efficiency than HDR and high accuracy. We also compared programmable transposase with a HITI variant in which we fused Cas9 to a catalytic dead version of PB, which may help in recruiting DNA to the insertion site as it has been recently suggested by a similar approach using the DNA binding domain of the

SB100 transposase. Programmable transposase presents twofold higher efficiency compared to alternative aided HITI methods (Fig. 18). With the aim to demonstrate programmable transposase activity in in vivo mouse models, we built mRNA versions of programmable transposase. We delivered programmable transposase to mice liver targeting Rosa26 genomic safe harbour using in vivo JetPEI reagent and observed high copy number of transgene compared to an endogenous gene (Fig. 11) and maintained transgene expression overtime (Fig. 20).

To sum up, we have coupled CRISPR molecular recognition and cleavage with DNA cut-and-paste activity of a modified PB to generate an efficient tool to perform precise and efficient gene delivery. This technology scales very well with payload size. We demonstrated its efficacy in Hek293T and mice liver. We envision programmable transposase technology as a generalized platform for therapeutic gene delivery for advanced therapies and other applications.

Results of Cas variants fused to hyPB

To further characterize the capacity of engineered hyPB to perform programmable transposition, we substituted the SpCas9 mudile of the programmable transposase tested, which other Cas proteins form different organisms with nuclease activity (namely SaCas9 of SEQ ID NO:72, cpf1 of SEQ ID NO:74, CasX of SEQ ID NO:75, and CjCas9 of SEQ ID NO:29). Specific gRNA targeting the region upstream of the split GFP reporter were designed and cloned for Hershey cell line transfection. (see Table 2 below) Targeted transposition was measured by means of GFP expression (Figure 12).

These results were confirmed in another line of experiments: we obtained good programmable insertion activity for CjCas9 and LbCpf1, while CasX did not achieve any programmable integration in our assay. Notably, SaCas9 had the highest levels of programmable insertion among the Cas proteins tested, with similar levels to SpCas9 fused to modified hyPB (Figure 19). Indels were determined for the different Cas proteins used and the three different gRNA designed for each protein by Ilumina NGS (Figure 21), shown for normalization purpose.

These positive results validate the engineered hyPB for programmable transposition to be useful for any sequence specific nuclease module.

Additional Results of Cas9 fused to a dimer of PB mutants

Given the nature of PB acting as dimers when performing transposition, we attempted to generate a fusion protein of Cas9 and hyPB R372A-K375A-D450N mutant. We compared the on target activity of these fusion to Cas9-PB mutants alone. We observed a better performance of the configuration Cas9-PB-PB; while the configuration PB-Cas9-PB did not outperform the Cas9-PB monomeric fusion (Figure 14). For the dimeric fusion to Cas9 we used a recorded version of hyPB R372A-K375A-D450N mutant to facilitate cloning and expression.

Interestingly, if the Cas9 fused to the dimeric hyPB R372A-K375A-D450N is the SaCas9 instead of SpCas9, the activity is further increased (Figure 24). The increased performance of SaCas9 over SpCas9 with the dimeric hyPB is consistent with the results obtained with monomeric hyPB (Figures 12 and 19).

Results of ZNF-PB mutants rescue

We wanted to further explore the role of the double-strand break (DSB) activity induced by close by (4 nucleotides) target sites of two gRNAs promoting single stranded cuts by SpCas9 nickase variant (D10A) in facilitating targeted integration, while lowering off target activity by means of non-inducing DSB in off target sites. We used a Zinc finger-PB fusion for directed localization of the transposon, by fusion with D450N mutant and R372A-K375A-D450N mutant and complemented it with two on-site single stranded brakes by an independent nickase Cas9, or single DSB by Cas9 nuclease.

Znf-PB fusion exhibited no or very low targeted insertion activity that was rescued when combined with introducing DSBs near the Znf binding site with a single or dual gRNA guided-cas9 either nuclease or nickase (Figure 13).

Results of Cas9-hyPB mutant variants

To further explore mutant combinations that could do programmable transposition with better efficiencies, several cycles of selection of cells were performed where GFP was reconstituted by programmable insertion of the split GFP reporter system. Interestingly we observed several combinations that out-performed the Cas9-hyPB R372A-K375A-D450N (Figure 15). Especially worth of mention is variant of hyPB fused to Cas9 that are mutated on hyPB at AA: A351-A372-A375-A388-N450-A465-A573-V589-G592-L594 (also identified as SEQ ID NO:2), several fold enrich in the positive cells population compared to R372A-K375A-D450N (SEQ ID NO:1); and also A245-A275-A277-A372-A465-V589 (SEQ ID NO:3) and A275-A325-A372-A560 (SEQ ID NO:4) to a lesser extent.

In another line of experiments, PiggyBac DNA library was produced by Twist Bioscience, cloned in fusion with cas9 into a lentiviral vector and transformed into stb4 competent cells, ensuring x100 variant complexity. Plasmids were purified by maxiprep and cotransfected with lentivirus packaging plasmids into Hek293T cells. Lentivirus was used to infect ½ GFP reporter cell line. Infected cells were transfected with the ½ GFP transposon and gRNA targeting AAVS1 sequence. GFP positive cells were selected by flow cytometry sorting and genomic DNA was extracted. PB was amplified from the extracted gDNA, recloned into lentiviral vector to restart a new cycle. Best performing programmable transposase variants were selected and transfected individually with AAVS1 gRNA and MC ½ GFP.

First, a random selection of 96 variants was performed and best performing variants were screened separately (Fig. 16). A summary of best PB amino acid variants for high on-target insertion confirms the importance of mutations D450N, R372A and K375A; but highlights other important residues which contribute to increased targeted efficiency (Fig. 17B). The six PB variants with best on-target efficiencies were selected (Fig. 17A). The individual on-target activities were significantly improved compared to FiCAT (Cas9-hyPB R372A-K375A-D450N) with the following variants: N347A-D450N; N347S-D450N-T560A-S573A-F594L; R202K-R275A-N347S-R372A-D450N-T560A-F594L; R275A-N347S-K375A-D450N-S592G; R275A-N347S-R372A-D450N-T560A-

F594L; and R275A-R277A-N347S-R372A-D450N-T560A-S564P-F594L (two-sided t-test).

This experiment was repeated and confirmed (Fig. 22A). We also produced lentiviruses expressing bulk variants of each cycle and infected reporter cell line correcting its titer by the PB variants CN, demonstrating a similar increase of on-target efficiency over cycles (Fig. 22B). Single mutants were isolated from bulk variants after 4 and 5 cycles of cas9_PB library enrichment. Mutants were tested separately by transfecting on-target reporter cell line with FiCAT mutant, gRNA tcr1 and ½ GFP MC transposon. Best FiCAT mutants are shown in comparison with FiCAT R372A_K375A_D450N (Fig. 23). The individual on-target activities were significantly improved compared to FiCAT (Cas9-hyPB R372A-K375A-D450N) with the following variants: R202K-R275A-N347S-R372A-D450N-T560A-F594L; R245A-N347S-R372A-D450N-T560A-S564P-S573A-S592G; R275A-N347S-R372A-D450N-T560A-F594L; N347A-D450N; R277A-G325A-N347A-K375A-D450N-T560A-S564P-S573A-S592G-F594L; N347S-D450N-T560A-S573A-F594L; V34M-R275A-G325A-N347S-S351A-R372A-K375A-D450N-T560A-S564P; G325A-N347S-K375A-D450N-S573A-M589V-S592G; S230N-R277A-N347S-K375A-D450N; T43I-R372A-K375A-A411T-D450N; G325A-N347S-S351A-K375A-D450N-S573A-M589V-S592G; Y177H-R275A-G325A-K375A-D450N-T560A-S564P-S592G.

The superiority of mutants R202K-R275A-N347S-R372A-D450N-T560A-F594L, R275A-R277A-N347S-R372A-D450N-T560A-S564P-F594L and R275A-N347S-R372A-D450N-T560A-F594L compared to the mutant R372A-K375A-D450N was further demonstrated in triple fusion proteins comprising a SpCas9 and two hyPB (Figure 29).

Results of Cas9 and hyPB non-covalent linking

In addition to the covalent binding of Cas9 with hyPB R372A-K375A-D450N through a linker, we user the MS2-MCP system to link Cas9 and the fusion protein consisting of MCP protein and hyPB R372A-K375A-D450N through a modified gRNA containing a tetraloop of MS2 sequence binding the MCP protein.

The combination of MCP-hyPB R372A-K375A-D450N fusion protein with Cas9 had an increased programmable insertion activity compared to Cas9- hyPB R372A-K375A-D450N fusion protein (Fig. 25A). In addition, we fused the MCP protein to other mutants of hyPB to perform programable transposition in combination with SpCas9. Both variants used (R202K-R275A-N347S-R372A-D450N-T560A-F594L, and R275A-N347S-R372A-D450N-T560A-F594L) outperformed R372A-K375A-D450N (Fig. 25B).

Results of Cas9 and hyPB decoupling for programable transposition

We also tried the performance of SpCas9 with hyPB R372A-K375A-D450N without a linker, nor the MS2-MCP system. We co-expressed in the same cells SpCas9 and hyPB R372A-K375A-D450N, and an increased programmable insertion activity was registered compared to Cas9- hyPB R372A-K375A-D450N fusion protein (Fig. 26A). We extended the number of hyPB mutant variant tested not fused to Cas9, but being expressed at the same time, and acting together to achieve the activity of programable transposition (Fig. 26B).

Results of co-expression of Cas9-hyPB and MCP-hyPB fusion proteins

We co-transfected hyPB R372A-K375A-D450N mutant fused to MCP protein, and hyPB mutants fused to SpCas9, in order to obtain a dimeric version of the fusion with one of the monomers being non-covalently linked. Several hyPB mutants fused to SpCas9 were compared for specific target integration (Fig. 27).

Results of co-expression of Cas9-hyPB and hyPB variants

In a similar manner, we co-transfected the SpCas9 hyPB R372A-K375A-D450N fusion protein and hyPB mutants independently, in order to obtain a dimeric version of the fusion with one of the monomers not linked (Fig. 28).

Methods

Cloning and plasmids. RFP transposon PB512-B for random insertion monitoring was purchased from System Biosciences Inc. hyPB vector was obtained from Wellcome Trust Sanger Institute (pCMV_hyPBase)[9]. Plasmid vector pCRTM-Blunt II-TOPO® was from

Invitrogen and cas9, ncas9 and SP-dcas9-VPR were obtained from Addgene (Addgene plasmid #41815, #41816, #63798). Finally, SB100X and pT4-HB were a kind gift from Dr. Zsuzsana Zizsvak. gRNAs were produced using The Zero Blunt TOPO PCR cloning kit (Invitrogen). with a gblock gene fragment (Integrated DNA Technologies) containing U6 promoter, 20 nt target site, gRNA scaffold and terminator. gRNA TRAC was designed and validated in the lab and gRNA aavs1 3 sequence was previously described[18].

Nuclease, nickase and dead cas9 fusions to hyPB and PB RFP ½ emGFP SMN1 transposon were performed by Golden Gate assembly using BspQI enzyme and standard methods.

Different mutations were introduced into hyPB sequence fused to cas9 (cas9_PB plasmid) by site directed mutagenesis following Quickchange Lightning mutagenesis kit's instructions (Agilent). Primers were designed with QuickChange Primer Design to achieve following mutations to the hyPB sequence: M194V, R245A, G325A, R372A, K375A, R376A, E377A, E380A, D450N, S564P (referring to SEQ ID NO:90-99). All plasmids are available upon request. PB ½ emGFP SMN1 was obtained by introducing the first half of emGFP sequence and SMN1 intron 6 sequence into piggyBac acceptor vector. pT4 SMN1 2/2 emGFP was obtained by adding a second half SMN1 intron 6 and partial emGFP in SB100X transposon vector. emGFP sequences containing SMN1 were obtained from DYP004reporter[19], a kind gift from Sri Kosuri.

Transposon and HDR templates of different sizes were generated by cloning a partial cDNA (NC_000006.[12]) fragment upstream of the split emGFP reporter system

Cell culture, transfection and electroporation. Hek293T cell line (Thermo Fisher Scientific) and C2C12 cell line (ATCC) were cultured at 37ºC in a 5% CO2 incubator with Dulbecco's modified eagle medium (DMEM), supplemented with high glucose (Gibco, Therm Fisher), 10% Fetal Bovine Serum (FBS), 2 mM glutamine and 100 U penicillin/0.1 mg/mL streptomycin. Jurkat cell line was cultured at 37ºC in a 5% CO2 incubator with Roswell Park Memorial Institute 1640 medium (RPMI) supplemented with Glutamax and HEPES (Gibco, Thermo Fisher) and 10% FBS. This cell line was a gift from Manel Juan (Hospital Clinic, Barcelona). Hek293T cell's transfection

experiments were performed using lipofectamine 3000 reagent following manufacturer's instructions or Polyethyleneimine (PEI, Thermo Fisher Scientific) at 1:3 DNA-PEI ratio in OptiMem. Cells were seeded the day before to achieve 70% confluency on transfection day (usually 290.000 cells in adherent p12 well plate). Plasmid DNA ratio was 1 transposase: 2,5 gRNA: 2,5 transposon or 1 Cas9 : 2,5 gRNA : 2,5 HDR template using either 0,076 pmols programmable transposase or Cas9 for p12 well plate.

emGFP splicing based reconstitution Assay. Hek293T cell line containing pT4 SMN1 2/2 emGFP was generated by PEI mediated transfection of SB100X and pT4 SMN1 2/2 emGFP DNA constructs, followed by single clone expansion and PCR genotyping (Supplementary Table 3). A positive clone was selected and expanded and used for subsequent assays. For emGFP reconstitution assay, programmable transposase, gRNA and transposon plasmids were transfected in a 1 programmable transposase: 2,5 gRNA : 2,5 transposon ratio using 0,076 pmol programmable transposase or hyPB and 0,19 pmols transposon and gRNA for a 12 wells plate. On-target insertion was measured 5 days post-transfection by emGFP fluorescence. Off-target transposition was measured 15 days post-transfection by RFP fluorescence. On target insertion was measured by cotransfecting PB variant, gRNA and PB ½SMN1 RFP emGFP (insert definitive construct name) expression plasmids and determining emGFP and RFP fluorescence after 14 days (set average n of days for episomal decay) by emGFP expression measured at (BD LSR Fortessa; BD Biosciences. Blue 488nm laser with 530/30 filter and Yellow Green 561nm laser with 610/20 filter)

Junction PCRs for insertion site sequencing Junction PCR was performed on emGFP sorted cells with BD FACSAria (Biosciences). Selected cells had on-target insertion of PB ½ emGFP SMN1 transposon targeting TRAC target site on reporter cell line. Genomic DNA was extracted using DNeasy Blood and tissue kit (Qiagen). Primers were designed by the 3' ITR of the transposon (forward) and targeting the intron of the 2/2 emGFP of the reporter cell line or the endogenous T cell receptor (TRAC) (reverse) (Supplementary Table 4).

Bioinformatics analysis of Guide-SEQ experiment. Illumina reads were clustered with usearch[20] and mapped to the reference using bwa-mem[21]. For on-target insertion

characterization, reads covering 5' and 3' junctions from the target insertion site were selected with Python scripting and Samtools[22]. Number of indels was obtained with CRISPR-GA[23]. For on-target and off-target experiments, clustered reads that mapped against the vector were selected and mapped against the reference genome using bwa-mem. Significance of the insertion peaks was assessed with macs2[24] algorithm.

Guide-seq library prep adapted to targeted insertion. An adapted Guide-seq[15] protocol implementation was performed by extracting genomic DNA using DNeasy Blood and tissue kit (Qiagen) and fragmented to 500bp fragments using Q800R3 Sonicator. End repair, A-tailing, and ligation of Y-adapter were performed using KAPA Hyper Prep Kit (KR0961 – v5.16) and 3ug of fragmented genomic DNA, followed by AMPure XP SPRI bead purification at 1X ratio. After adapter ligation, each sample was split in two and amplified with GSP5' or GSP3' to capture 5' and 3' junctions, respectively.To capture 5' and 3' transposon-genome junctions, two nested PCRs were performed using KAPA HiFi DNA Polymerase following manufacturer protocol: PCR1 with P5_1 and PB_5_GSP1 or PB_3_GSP1 in a 25ul final volume; and PCR2 with P5_2 PB_5_GSP2 or PB_3_GSP2 in a 25ul final volume. 5' and 3' PCR products were purified with AMPure XP SPRI bead purification at 1X ratio, mixed in equimolar ratio and sequenced with Illumina Miseq Reagent Kit V2 - 500 cycle (2 x 250 bp paired end). 3 ul of 100 μM custom primers index 1 and Read 2 were added to the sequencing reaction.

in vivo targeted insertion to mice liver. Animal experimentation procedures were approved by the Animal Experimentation Ethic Committee of Barcelona Biomedical Research Park. C57BL/6J, 8-10 weeks old were used for this study. The animals were purchased from Jackson Laboratories, male and female were used without distinction. programmable transposase mRNA was produced with RiboMAX Large Scale RNA Production Systems-T7 (Promega) following manufacturer's instructions. Rosa26 gRNA[25] was purchased from IDT. programmable transposase mRNA, gRNA targeting Rosa26 and PB512-B transposon were injected via retro-orbital in a 1:1:2,5 ratio. A total of 60 ug of nucleic acids were complexed to; In vivo-JetPEI (Polyplus transfection) at NP ratio 7. Animals were euthanized 10 days after-injection and liver was isolated and homogenized. Genomic DNA was extracted from liver samples with DNeasy Blood and

tissue kit (Qiagen) Transposon relative Copy number to Tfrc endogenous gene was obtained by qPCR (primers listed in Supplementary Table 1). Imaging of luciferase expression was performed at different timepoints after FiCAT-gRNA-transposon or transposon control administration with IVIS spectrum imaging system (Caliper Life Sciences). Images were taken 5 min after intraperitoneal injection of D-Luciferin potassium salt (Gold Biotechnology) according to the manufacturer's instructions.

PB structural modelling. A 3D structure of the Trichoplusia ni piggyBac transposase protein was obtained by Robetta Web protein structure prediction server (http://robetta.bakerlab.org). The core domain (131-550aa) was predicted by Rosetta Comparative Modelling method that is based on Monte Carlo algorithm with embedded Cartesian-space minimization and all-atom optimization[26]. The tertiary structure fold was analysed and validated with SPServer and ProSa-Web knowledge-based methods (Supplementary Fig. 2). Secondary structure was analysed with PSIPRED and HHPred machine-learning based methods. PB's core was then modelled for refinements with PyMOL by comparative protein modelling methods. The refinement process was guided by the superimposition of the piggyBac model with Cryo-EM HIV-1 Strand Transfer Complex Intasome (PDB ID: 5U1C) consisting of the HIV integrase tetramer bound to viral DNA and target host DNA and X-ray diffraction Tn5 transposase complex structure (PDB ID: 1MUS27). Strand-transferring DNA and donor DNA were extrapolated from the superimpositions of HIV-1 Intasome and Tn5 respectively. The nucleotides in the interface in contact with the protein were analyzed with X3DNA as double-strand DNA. We used statistical potentials to score the interaction between protein and DNA and generate a theoretical PWM[28]. The theoretic PWM is obtained by testing all potential double-strand DNA sequences in the interface, ranking them with the statistical potentials and selecting the top to make a multiple sequence alignment. During the submission of this manuscript a cryo-EM structure became available, which shows important agreement with modelling performed[29]. Cryo-EM structure of piggyBac transposase strand transfer complex (PDB ID: 6X67) confirmed the general fold of the model and the domains we hypothesized were responsible for the contact with donor and target DNA.

Table 2 : Cas variants gRNA

Cas Variant	Sequence	PAM
SaCas9	TATGTACACTTCTGACCCAC (SEQ ID NO: 113)	TGGAAT
	GCCTTTAAGCTTGATATCCA (SEQ ID NO: 114)	TGGAAT
	GTATCACAATTCCAGTGGGT (SEQ ID NO: 115)	CAGAA
	GGACAGGATCGGCATAACCG (SEQ ID NO: 116)	GTGAAT
	GTGCTCGGGGCCACTAGGGA (SEQ ID NO: 117)	CAGGAT
Cpf1	ACTTATAATTCACTGTATCA (SEQ ID NO: 118)	TTTC
	AGCTTGATATCCATGGAATT (SEQ ID NO: 119)	TTTA
	TGCTCGGGGCCACTAGGGAC (SEQ ID NO: 120)	TTTG
	CTTTTGTAAAACTTTATGGT (SEQ ID NO: 121)	TTTA
	CAAAAGTAAATAGCCCGGCT (SEQ ID NO: 122)	TTTA
CjCas9	GCCGATCCTGTCCCTAGTGGCC (SEQ ID NO: 123)	CCGAGCAC
	ACAATTCCAGTGGGTCAGAAGT (SEQ ID NO: 124)	GTACATAC
	ACACTTCTGACCCACTGGAAT (SEQ ID NO: 125)	TGTGATAC
	GAATTCCATGGATATCAAGCTT (SEQ ID NO: 126)	AAAGGCAC
	AATTCCAGTGGGTCAGAAGTGT (SEQ ID NO: 127)	ACATACAC
CasX	TCAAGCGCGTGTATGTACAC (SEQ ID NO: 128)	TTCT
	GGATCGGCATAACCGGTGAA (SEQ ID NO: 129)	TTCC
	TAGACATGAGGTCTATGGAC (SEQ ID NO: 130)	TTCA
	TAAGCTTGATATCCATGGAA (SEQ ID NO: 131)	TTCA
	TATAATTCACTGTATCACAA (SEQ ID NO: 132)	TTCC

REFERENCES

1. Porteus, M.H. & Carroll, D. *Nat. Biotechnol.* **23**, 967–973 (2005).
2. Sander, J.D. & Joung, J.K. *Nat. Biotechnol.* **32**, 347–355 (2014).
3. Rees, H.A. & Liu, D.R. *Nat. Rev. Genet.* doi:10.1038/s41576-018-0059-1
4. Anzalone, A.V. et al. *Nature* **576**, 149–157 (2019).
5. He, X. et al. *Nucleic Acids Res.* **44**, e85 (2016).
6. Suzuki, K. et al. *Nature* **540**, 144–149 (2016).
7. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. & Sternberg, S.H. *Nature* 1 (2019).
8. Strecker, J. et al. *Science* (2019).doi:10.1126/science.aax9181
9. Yusa, K., Zhou, L., Li, M.A., Bradley, A. & Craig, N.L. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1531–1536 (2011).
10. Hew, B.E., Sato, R., Mauro, D., Stoytchev, I. & Owens, J.B. *Synth. Biol.* **4**, ysz018 (2019).
11. Kovač, A. et al. *Elife* **9**, (2020).
12. Loperfido, M. et al. *Nucleic Acids Research* **44**, 744–760 (2016).
13. Passos, D.O. et al. *Science* **355**, 89–92 (2017).
14. Li, X. et al.doi:10.1073/pnas.1305987110
15. Morellet, N. et al. *Nucleic Acids Res.* **46**, 2660–2677 (2018).
16. Li, M.A. et al. *Mol. Cell. Biol.* **33**, 1317–1330 (2013).
17. Tsai, S.Q. et al. *Nat. Biotechnol.* **33**, 187–197 (2015).
18. Mali, P. et al. *Science* **339**, 823–826 (2013).
19. Cheung, R. et al. *Molecular Cell* **73**, 183–194.e8 (2019).
20. Edgar, R.C. *Bioinformatics* **26**, 2460–2461 (2010).
21. Li, H. *arXiv [q-bio.GN]* (2013).at <http://arxiv.org/abs/1303.3997>
22. Li, H. et al. *Bioinformatics* **25**, 2078–2079 (2009).
23. Guell, M., Yang, L. & Church, G.M. *Bioinformatics* **30**, 2968–2970 (2014).
24. Gaspar, J.M.496521 (2018).doi:10.1101/496521
25. Chu, V.T. et al. *BMC Biotechnol.* **16**, 4 (2016).
26. Fu, D.Y. (2018) at <https://etd.library.vanderbilt.edu/available/etd-08012018-164524/unrestricted/DarwinYFu_Thesis_Submit.pdf>
27. Steiniger-White, M., Rayment, I. & Reznikoff, W.S. *Curr. Opin. Struct. Biol.* **14**, 50–57 (2004).
28. Meseguer, A. et al. *NAR Genom Bioinform* **2**, (2020).
29. Chen, Q. et al. *Nat. Commun.* **11**, 3446 (2020).

CLAIMS

1. A composition comprising

(i) a first protein comprising or consisting of a site-specific DNA binding protein capable of binding and cleaving a target nucleic acid sequence; or a nucleic acid construct encoding said first protein; and

(ii) a second protein comprising or consisting of a transposase; or a nucleic acid construct encoding said second protein;

wherein said transposase is a modified hyperactive PiggyBac, comprising one or more amino acid mutations as compared to hyperactive PiggyBac of SEQ ID NO: 9.

2. The composition according to claim **1**, wherein the first protein and the second protein are fused together to form a fusion protein, optionally through a linker.

3. The composition according to claim **2**, wherein the first protein is fused to the C-terminal end of the second protein, optionally through a linker.

4. The composition according to any one of claims **1** to **3**, wherein said transposase is a modified hyperactive PiggyBac, comprising one or more amino acid mutations to increase excision activity as compared to unmodified hyperactive PiggyBac, and/or one or more amino acid mutations to decrease DNA binding activity as compared to unmodified hyperactive PiggyBac.

5. The composition according to any one of claims **1** to **4**, wherein said one or more amino acid mutations do not consist of R372A, K375A, and D450N.

6. The composition according to any one of claims **1** to **5**, wherein said one or more amino acid mutations are selected among the amino acid substitutions which increase excision activity at position of M194, D450, T560, S564 S573, S592 or F594, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions M194V and/or D450N.

7. The composition according to any one of claims **1** to **6**, wherein said one or more amino acid mutations are selected among the amino acid substitutions which increase excision activity at position of M194 or D450, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions M194V and/or D450N.

8. The composition according to any one of claims **1** to **7**, wherein said one or more amino acid mutations are selected among the amino acid substitutions which decrease DNA binding activity at position R275, R277, R347, R372, K375, R376, E377, and/or E380, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions R275A, R277, R347S, R372A, K375A, R376A, E377A, and/or E380A.

9. The composition according to any one of claims **1** to **8**, wherein said one or more amino acid mutations are selected among the amino acid substitutions which decrease DNA binding activity at position R372, K375, R376, E377, and/or E380, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9, preferably selected among the amino acid substitutions R372A, K375A, R376A, E377A, and/or E380A.

10. The composition according to any one of claims **1** to **9**, wherein the modified hyperactive PiggyBac includes the double mutations N347S and D450N, said position number corresponding to the amino acid number of unmodified hyperactive PiggyBac of SEQ ID NO: 9.

11. The composition according to any one of claims **1** to **10**, wherein the modified hyperactive PiggyBac mutation comprises one of the following amino acid substitution or combination of amino acid substitutions: R372A/K375A/R376A/D450N, K375A/R376A/E377A/E380A/D450N, R372A/K375A/R376A/E377A/E380A/D450N, M194V, R376A, E377A, E380A, M194V/R372A/K375A,

S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/S592G/F594L, R245A/R275A/R277A/R372A/W465A/M589V, R275A/325A/R372A/T560A, N347A/D450N, N347S/D450N/T560A/S573A/F594L, R202K/R275A/N347S/R372A/D450N/T560A/F594L, R275A/N347S/K375A/D450N/S592G, R275A/N347S/R372A/D450N/T560A/F594L, R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L, R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G, R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L, V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P, G325A/N347S/K375A/D450N/S573A/M589V/S592G, S230N/R277A/N347S/K375A/D450N, T43I/R372A/K375A/A411T/D450N, G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G, Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G; the position number corresponding to the amino acid number of the hyperactive PiggyBac of SEQ ID NO: 9, typically said modified transposase has an amino acid sequence selected among any of SEQ ID NO: 2-8, 10-18 and 135-149.

12. The composition according to any one of claims **1** to **11**, further comprising a third protein comprising or consisting of a second transposase; or a nucleic acid construct encoding said third protein; wherein said second transposase is either an hyperactive PiggyBac with SEQ ID NO: 9, or a modified hyperactive PiggyBac with comprising one or more amino acid mutations as compared to the hyperactive PiggyBac with SEQ ID NO: 9.

13. The composition according to claim **12**, wherein the first, second and third proteins are fused together to form a triple fusion protein, optionally through a linker.

14. The composition according to any one of claims **1** to **13**, wherein the first protein comprises or consists of an RNA-guided nuclease or nickase, or a zinc finger nuclease.

15. The composition according to any one of claims **1** to **14**, wherein said first protein is a nuclease protein comprising an active DNA cleavage domain and a guide RNA binding domain and having at least 80%, 90%, 95%, 99% or at least 100% identity to a *Streptococcus pyogenes* Cas9 (SpCas9) of SEQ ID NO: 31, *Staphylococcus aureus* Cas9 (SaCas9) of SEQ ID NO: 72, Cpf1 of SEQ ID NO: 74, *Campylobacter jejuni* Cas9 (CjCas9) of SEQ ID NO: 29, *Streptococcus pyogenes* Cas9 nickase (nCas9) of SEQ ID NO: 70, CasX of SEQ ID NO: 75, or *Staphylococcus aureus* Cas9 nickase of SEQ ID NO: 76;
preferably wherein said first protein is a Cas9 protein selected from the group consisting of a *Staphylococcus aureus* Cas9 (SaCas9) of SEQ ID NO: 72 and *Streptococcus pyogenes* Cas9 (SpCas9) of SEQ ID NO: 31.

16. The composition according to any one of claims **1** to **15**, further comprising a guide RNA, and an exogenous nucleic acid for insertion in a genome.

17. The composition according to any one of claim **16**, wherein the transposase is fused to an RNA-binding protein capable of binding to at least one specific RNA sequence comprised in the guide RNA;
optionally wherein said RNA-binding protein is an MS2 bacteriophage coat protein (MCP) and wherein the guide RNA comprises a MS2 RNA tetraloop binding sequence, preferably sharing at least 75% identity with SEQ ID NO: 153.

18. The composition according to claim **16** or **17**, wherein said exogenous nucleic acid is a large DNA fragment, typically having a size between 5 kb and 25 kb, and more preferably between 8 kb and 20 kb.

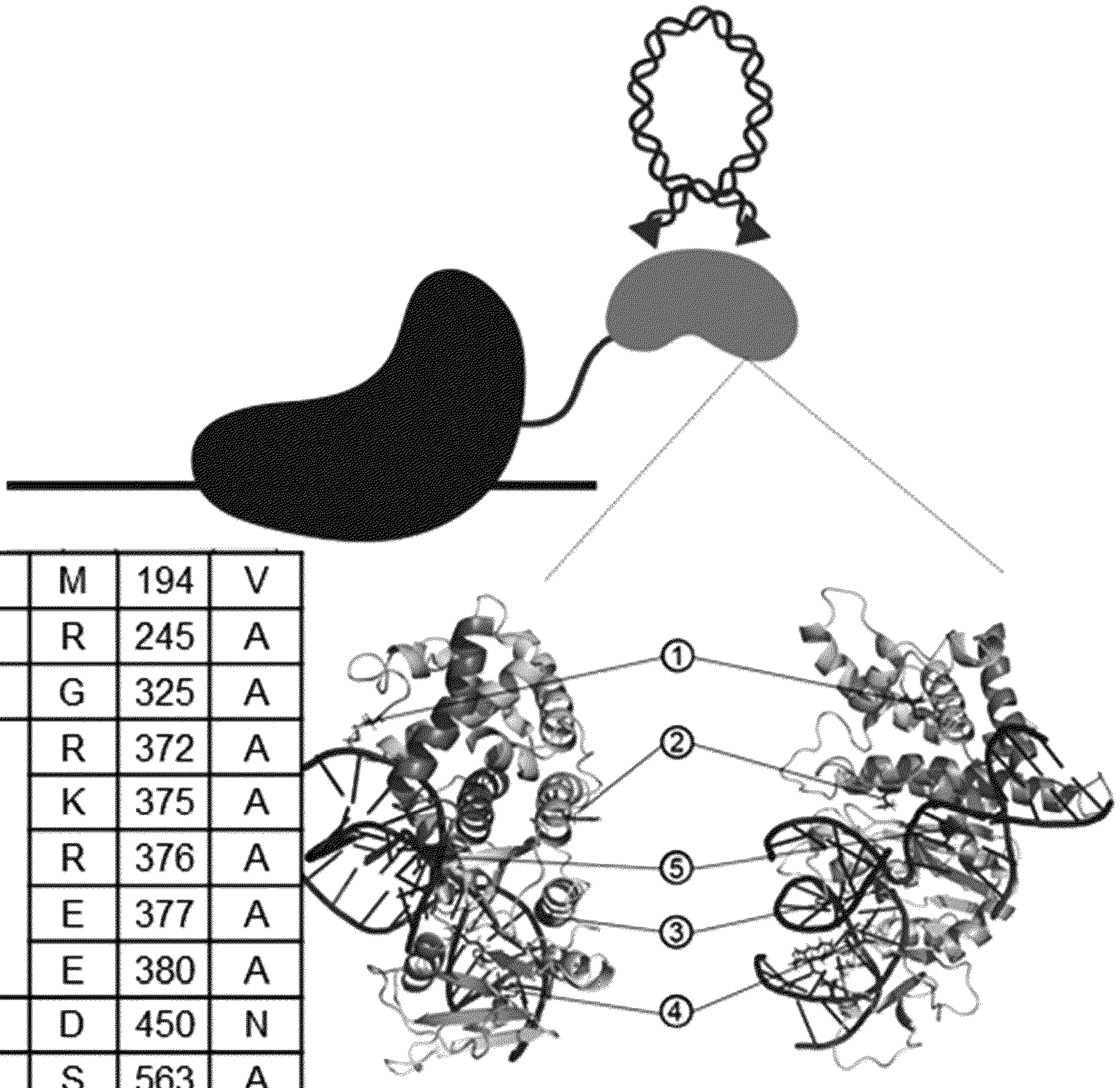
19. The composition according to any one of claims **1** to **18**, wherein said composition is comprised in a nanoparticle.

20. A nucleic acid encoding a fusion protein as defined in any one of claims **2** to **17**, typically a messenger RNA (mRNA).

21. An *in vitro* method for site specific integration of an exogenous nucleic acid sequence into the genome of a cell, the method comprising delivering to the cell the

composition according to any one of claims **1** to **19**, a guide RNA, and the exogenous nucleic acid.

22. The composition according to any one of claims **1** to **19**, a guide RNA, and an exogenous nucleic acid, for use in the treatment of a disease, by site-specific integration of the exogenous nucleic acid sequence into the genome of a cell.



1	M	194	V
2	R	245	A
3	G	325	A
4	R	372	A
	K	375	A
	R	376	A
	E	377	A
	E	380	A
5	D	450	N
	S	563	A

FIG. 1

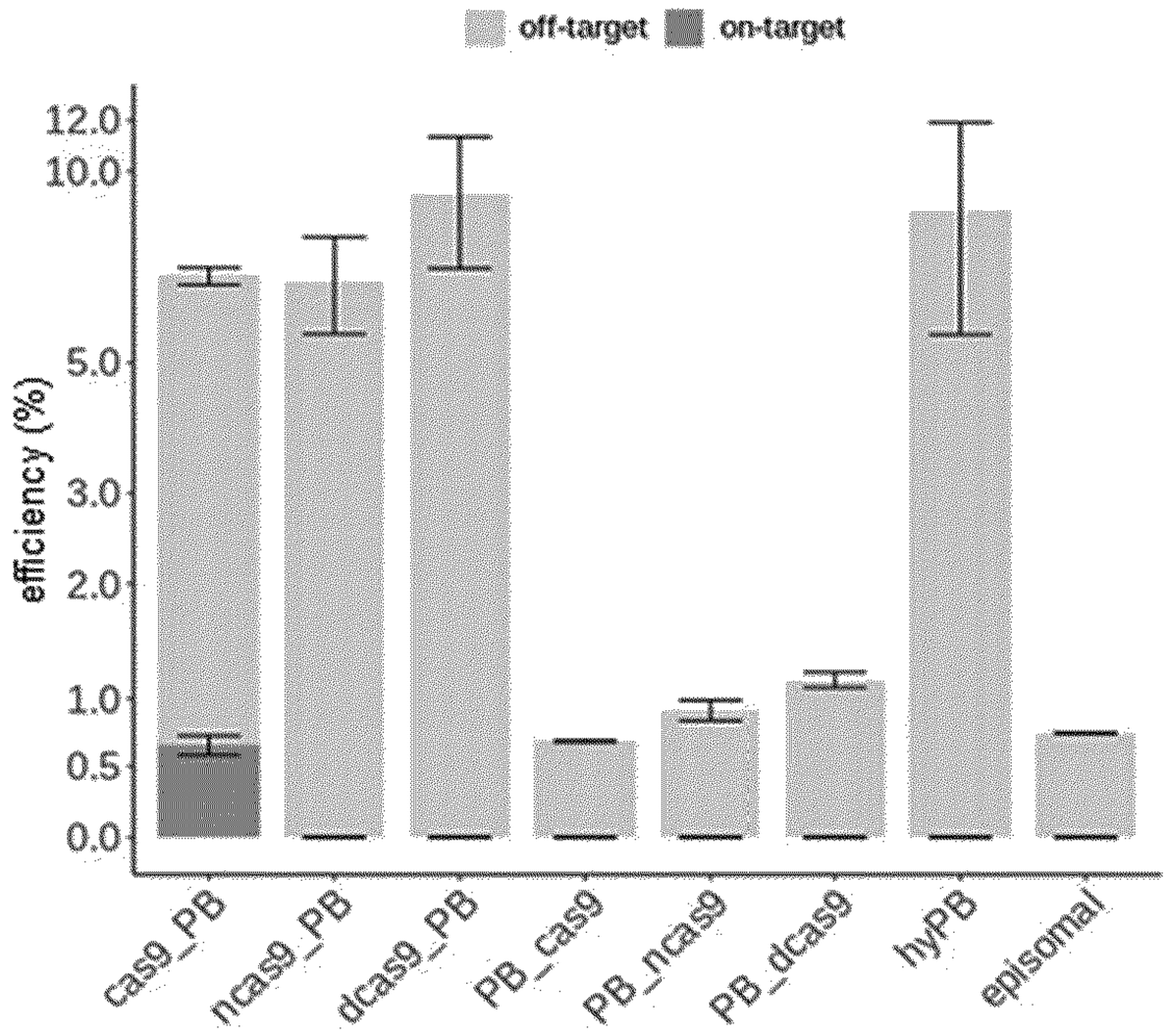

off-target
on-target
12.0
10.0
5.0
3.0
2.0
1.0
0.5
0.0
efficiency (%)
cas9_PB
ncas9_PB
dcas9_PB
PB_cas9
PB_ncas9
PB_dcas9
hyPB
episomal

FIG. 2A

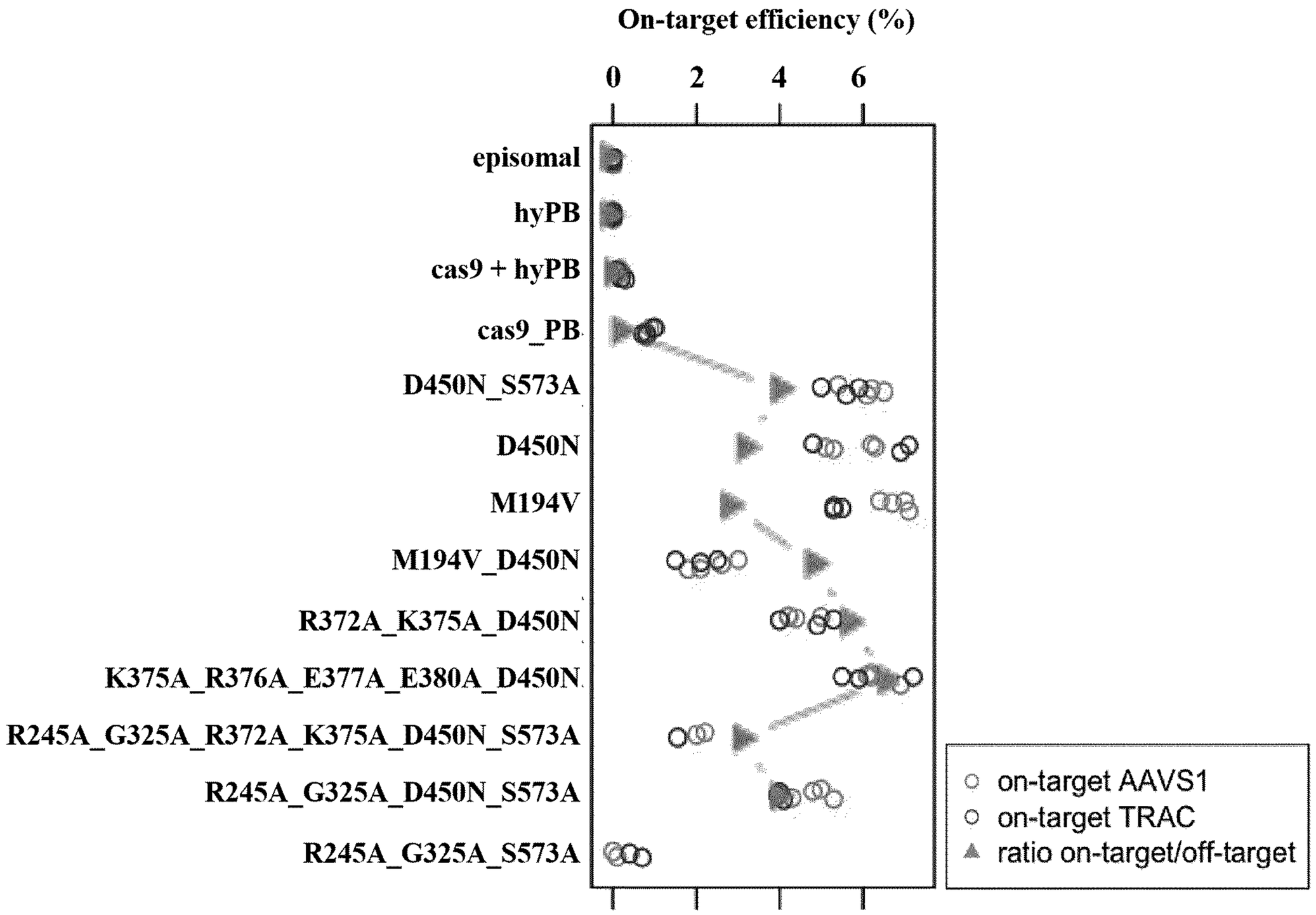

On-target efficiency (%)
0
2
4
6
episomal
hyPB
cas9 + hyPB
cas9_PB
D450N_S573A
D450N
M194V
M194V_D450N
R372A_K375A_D450N
K375A_R376A_E377A_E380A_D450N
R245A_G325A_R372A_K375A_D450N_S573A
R245A_G325A_D450N_S573A
R245A_G325A_S573A
on-target AAVS1
on-target TRAC
ratio on-target/off-target

FIG. 2B

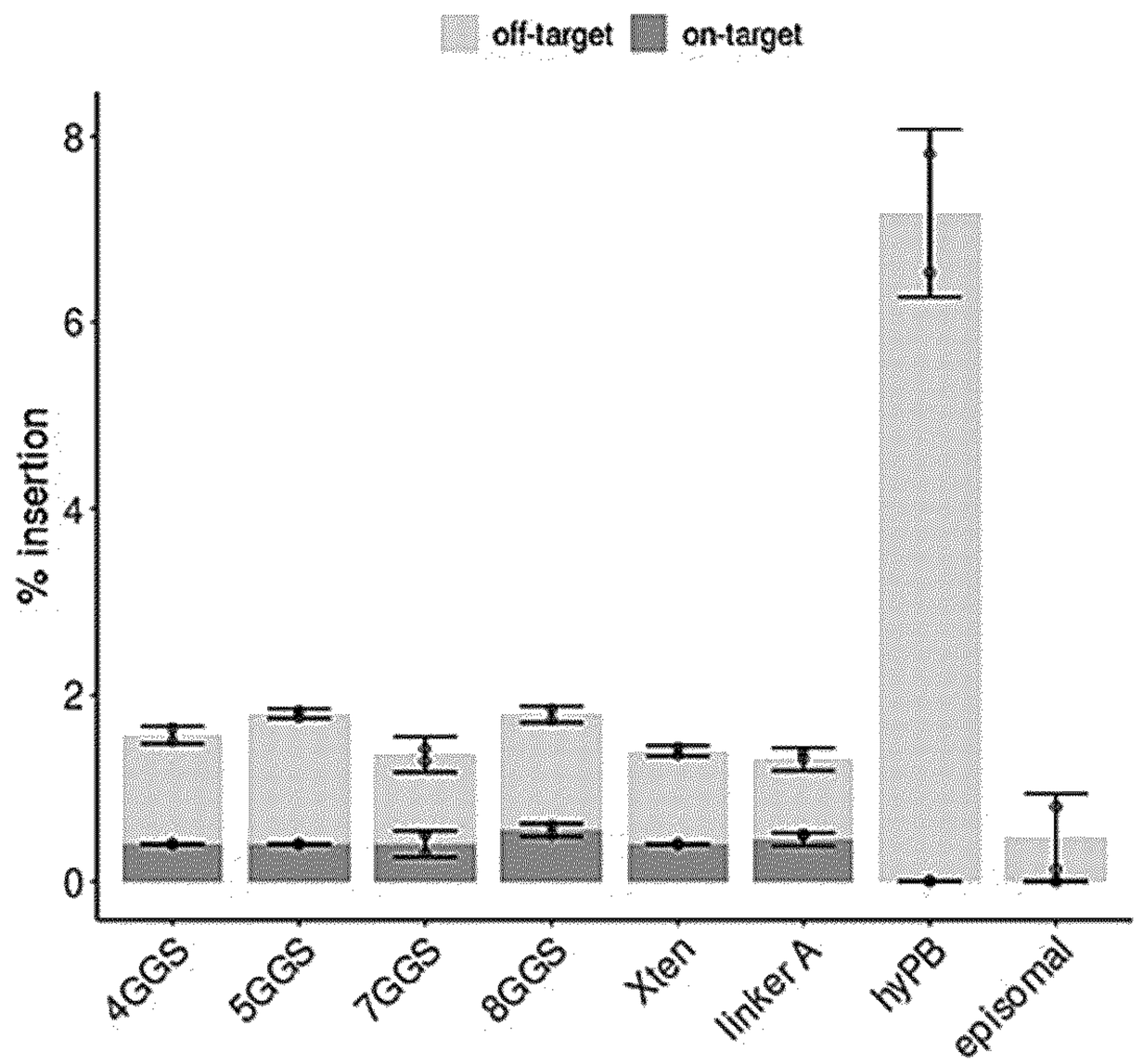

off-target
on-target
% insertion
0
2
4
6
8
4GGS
5GGS
7GGS
8GGS
Xten
linker A
hyPB
episomal

FIG. 2C

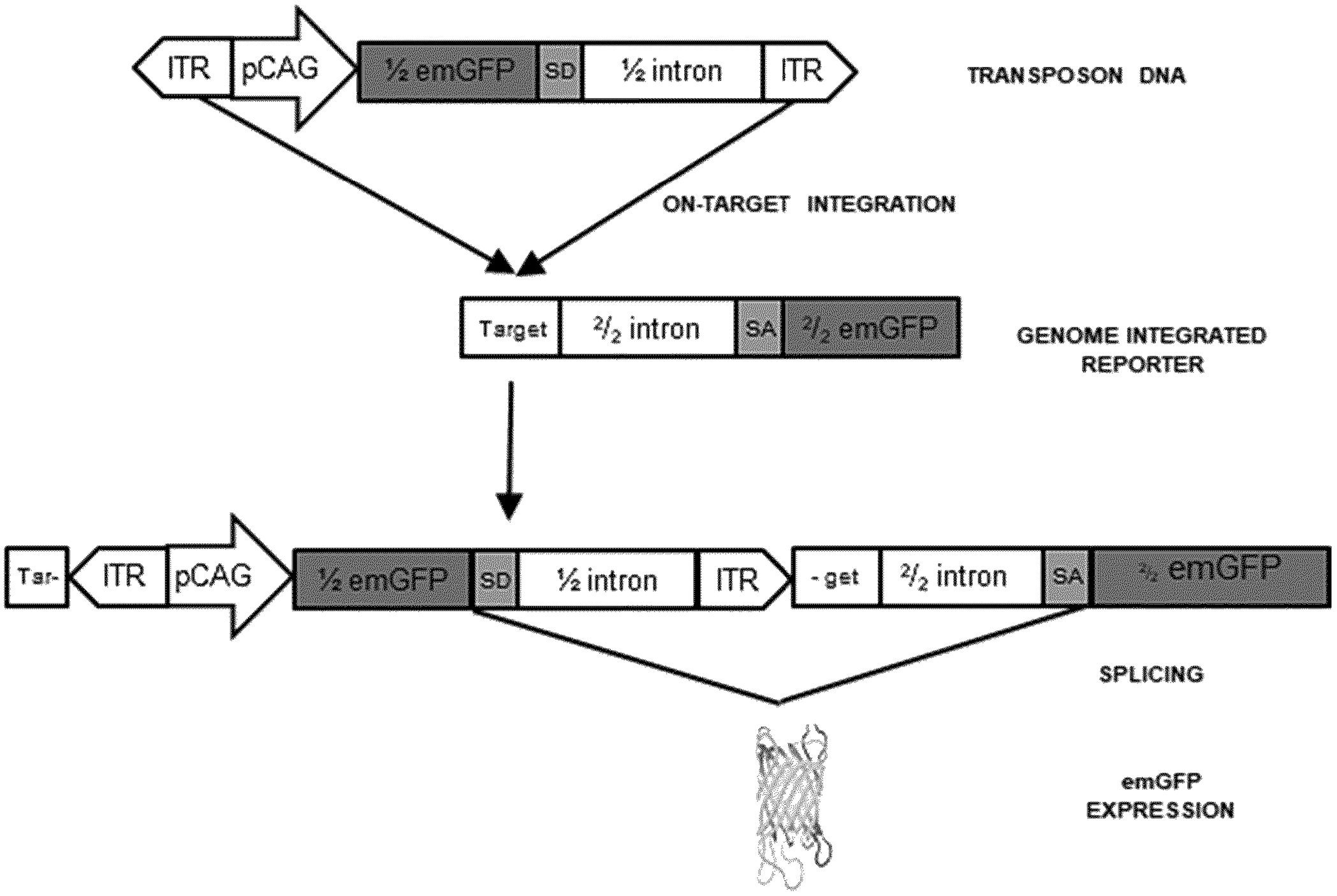

ITR
pCAG
½ emGFP
SD
½ intron
ITR
TRANSPOSON DNA
ON-TARGET INTEGRATION
Target
2/2 intron
SA
2/2 emGFP
GENOME INTEGRATED REPORTER
Tar-
ITR
pCAG
½ emGFP
SD
½ intron
ITR
-get
2/2 intron
SA
2/2 emGFP
SPLICING
emGFP EXPRESSION

FIG. 3

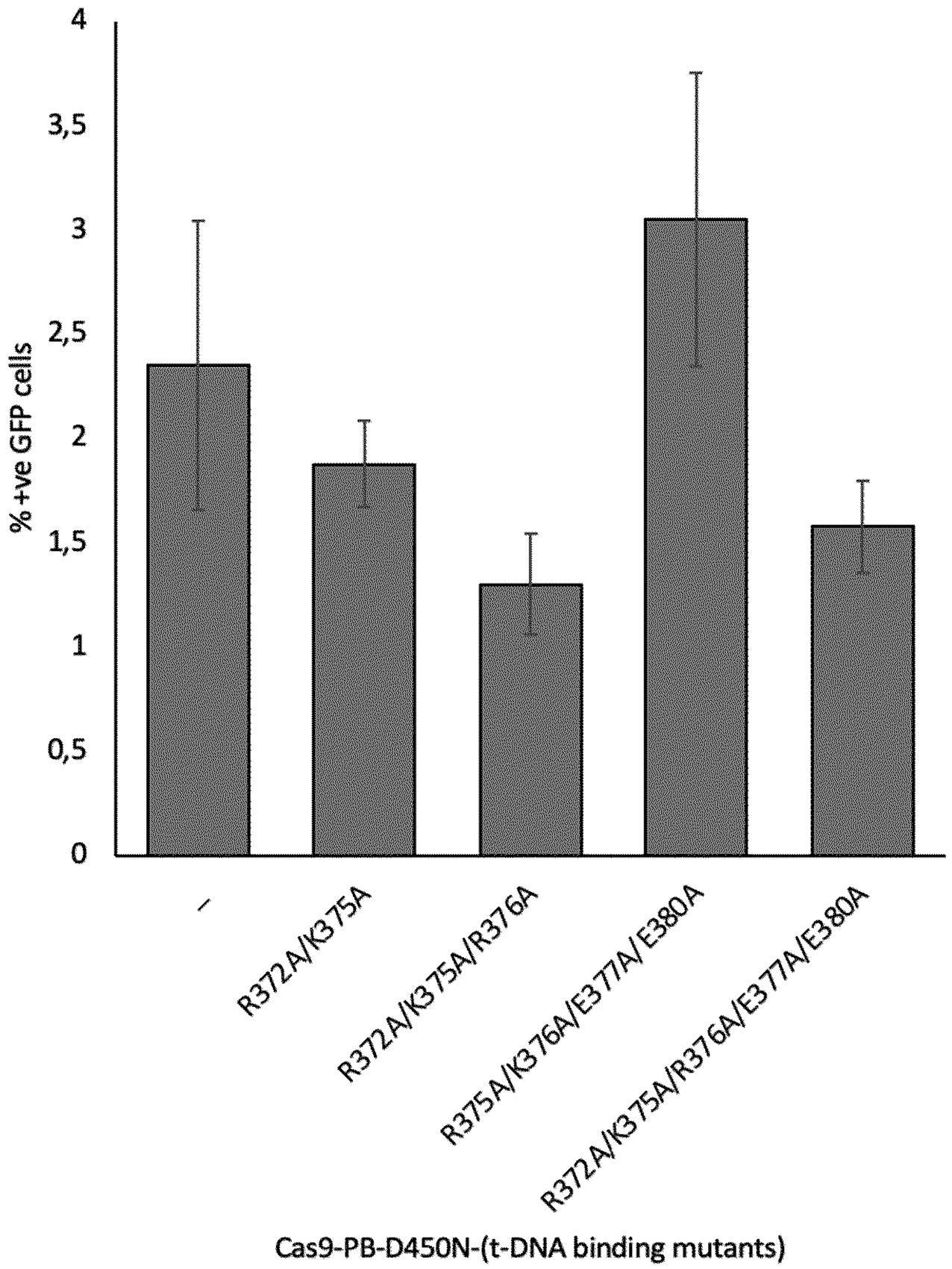

4
3,5
3
2,5
2
1,5
1
0,5
0
% +ve GFP cells
–
R372A/K375A
R372A/K375A/R376A
R375A/K376A/E377A/E380A
R372A/K375A/R376A/E377A/E380A
Cas9-PB-D450N-(t-DNA binding mutants)

FIG. 4A

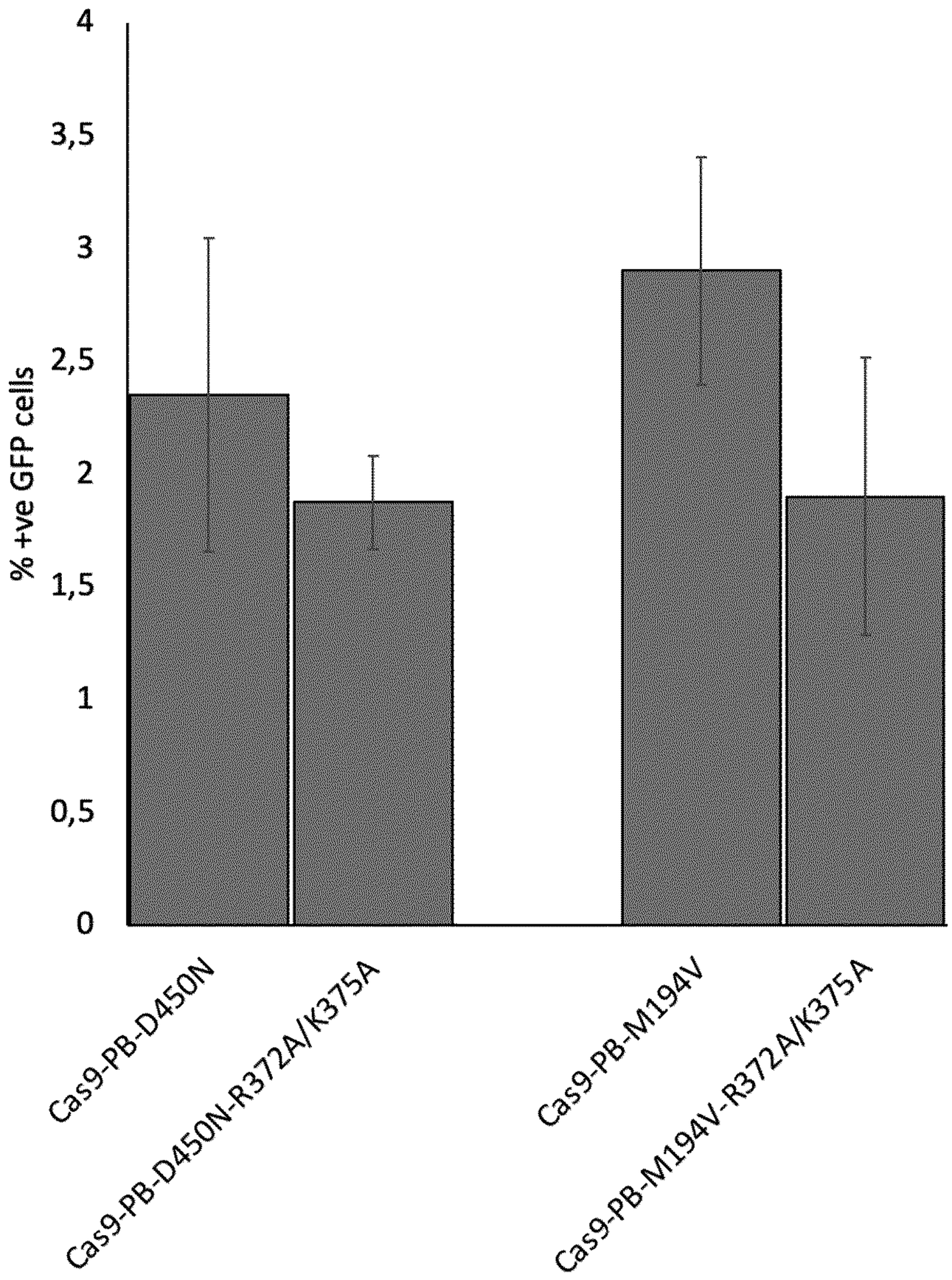

4
3,5
3
2,5
2
1,5
1
0,5
0
% +ve GFP cells
Cas9-PB-D450N
Cas9-PB-D450N-R372A/K375A
Cas9-PB-M194V
Cas9-PB-M194V-R372A/K375A

FIG. 4B

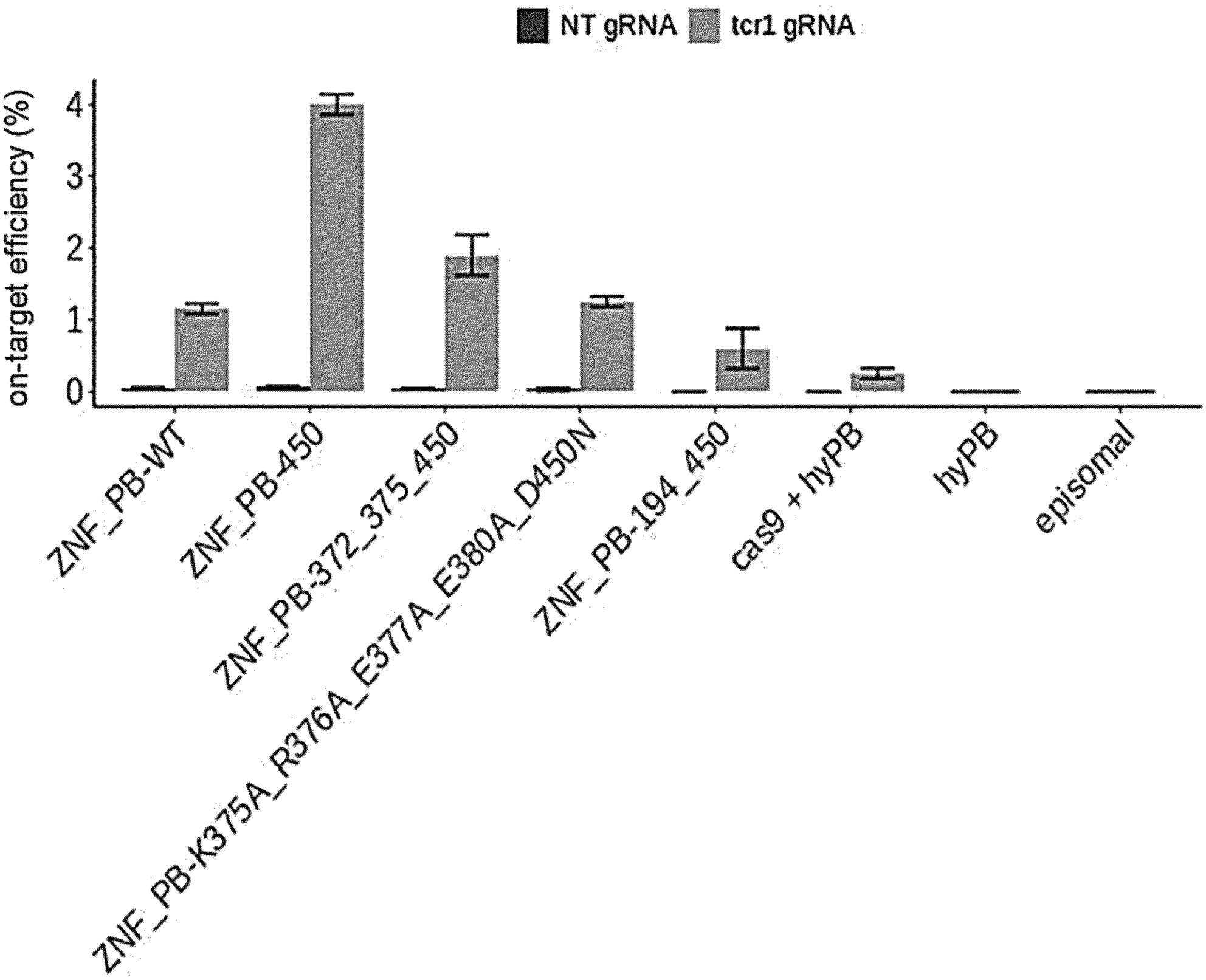

NT gRNA
tcr1 gRNA
on-target efficiency (%)
0
1
2
3
4
ZNF_PB-WT
ZNF_PB-450
ZNF_PB-372_375_450
R376A_E377A_E380A_D450N
ZNF_PB-194_450
cas9 + hyPB
hyPB
episomal
ZNF_PB-K375A_

FIG. 5

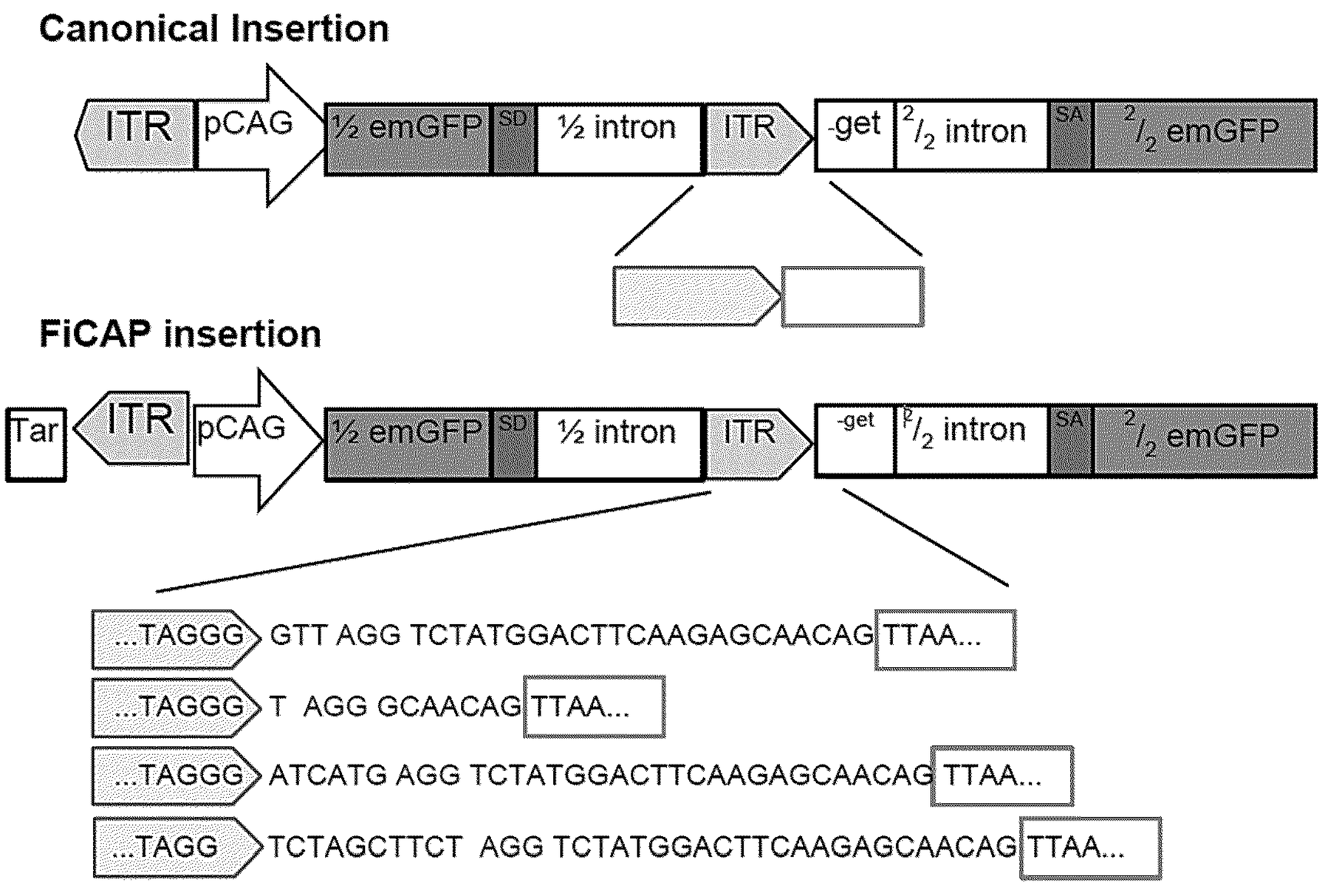

Canonical Insertion
ITR
pCAG
½ emGFP
SD
½ intron
ITR
-get
2/2 intron
SA
2/2 emGFP
FiCAP insertion
Tar
ITR
pCAG
½ emGFP
SD
½ intron
ITR
-get
2/2 intron
SA
2/2 emGFP
...TAGGG GTT AGG TCTATGGACTTCAAGAGCAACAG TTAA...
...TAGGG T AGG GCAACAG TTAA...
...TAGGG ATCATG AGG TCTATGGACTTCAAGAGCAACAG TTAA...
...TAGG TCTAGCTTCT AGG TCTATGGACTTCAAGAGCAACAG TTAA...

FIG. 6

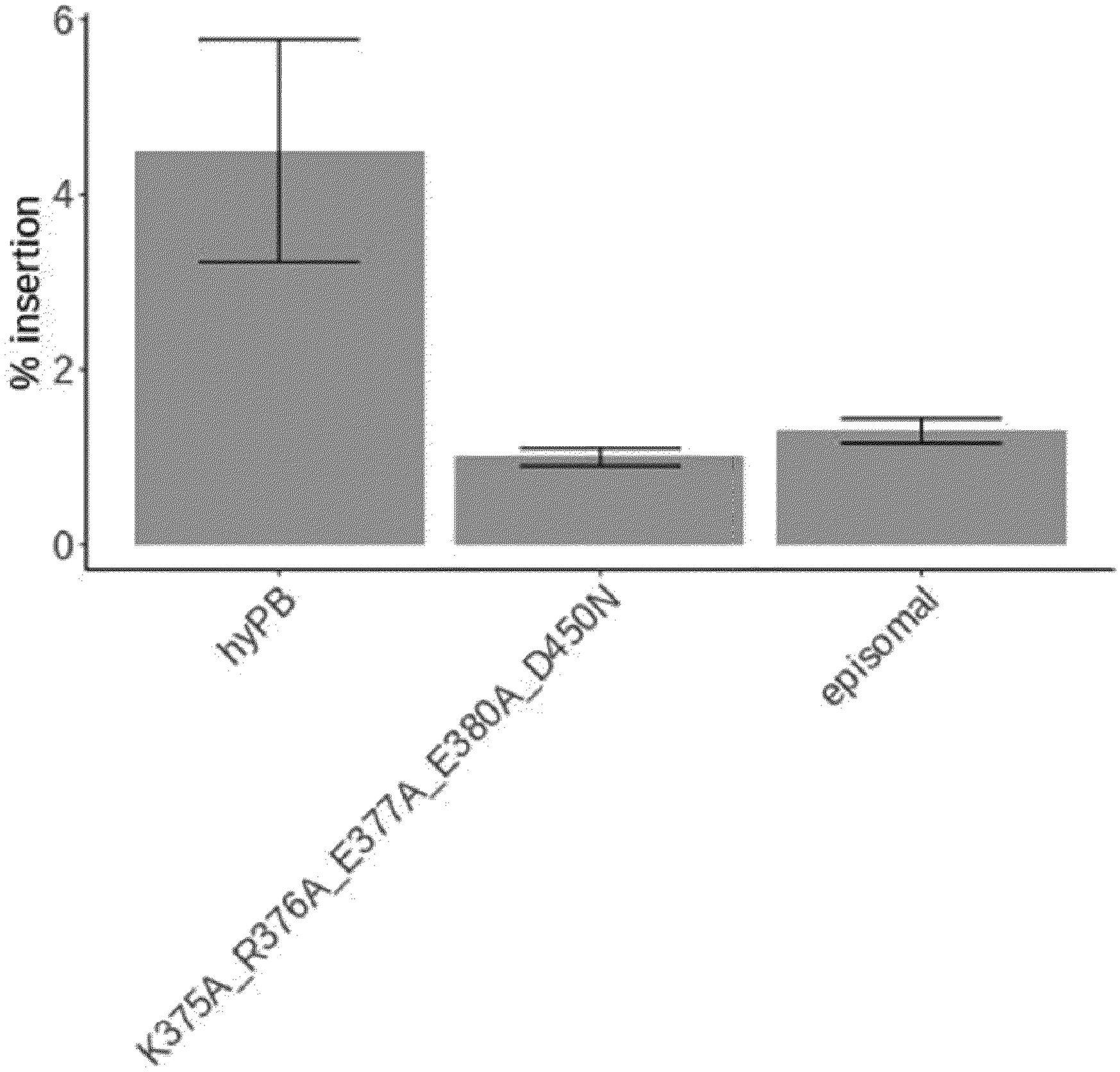

% insertion
6
4
2
0
hyPB
E377A_E380A_D450N
episomal
K375A_R376A_

FIG. 7

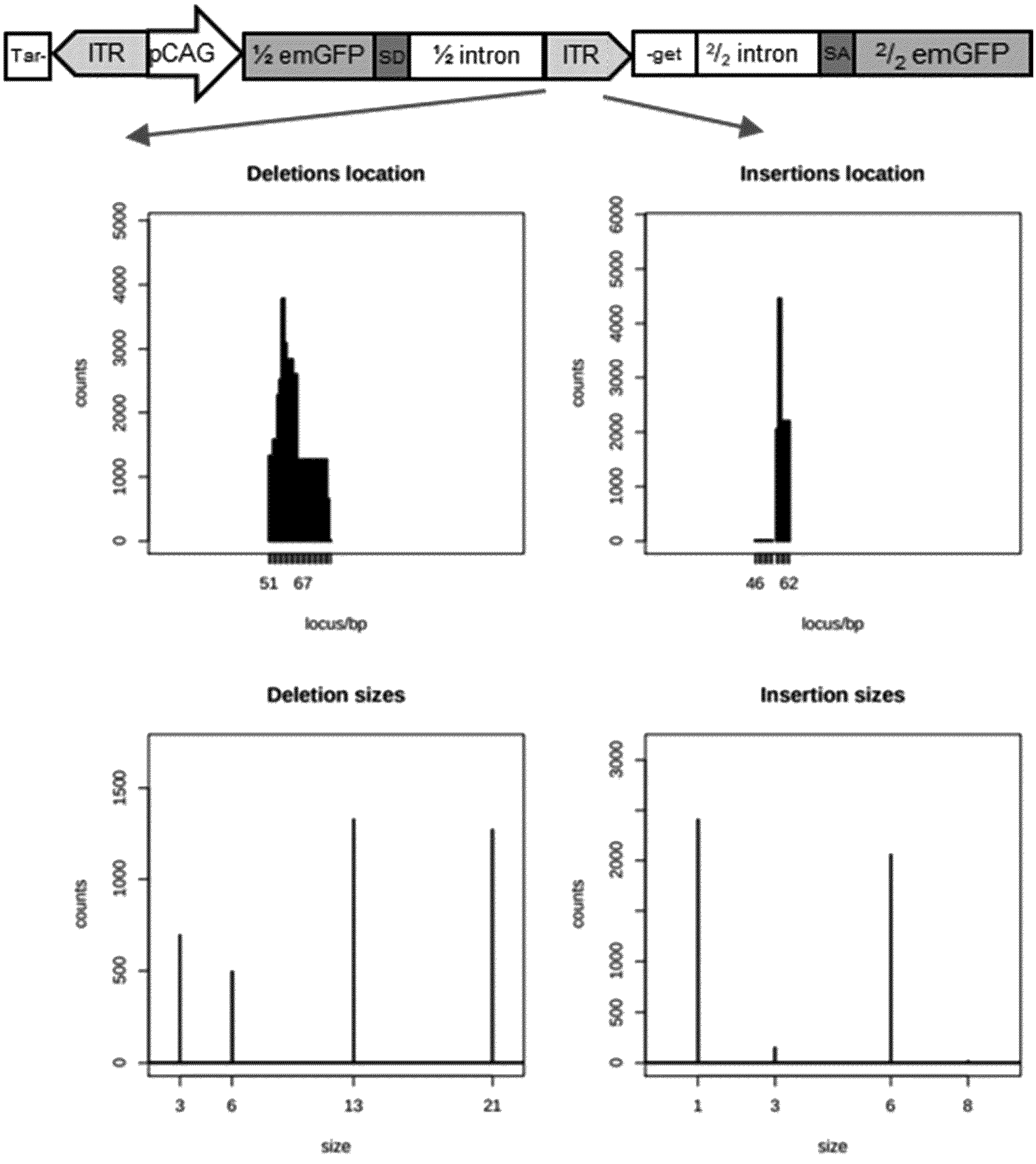
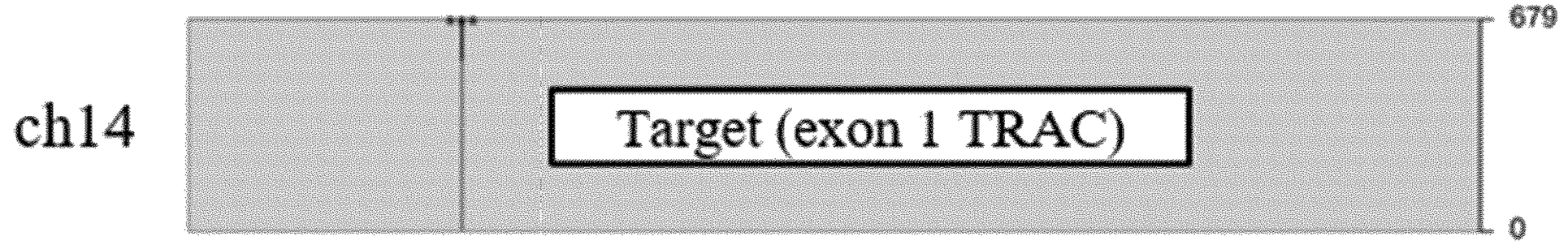

Ter-
ITR
pCAG
½ emGFP
SD
½ intron
ITR
-get
2/2 intron
SA
2/2 emGFP
Deletions location
Insertions location
counts
locus/bp
51
67
46
62
Deletion sizes
Insertion sizes
size
3
6
13
21
1
3
6
8

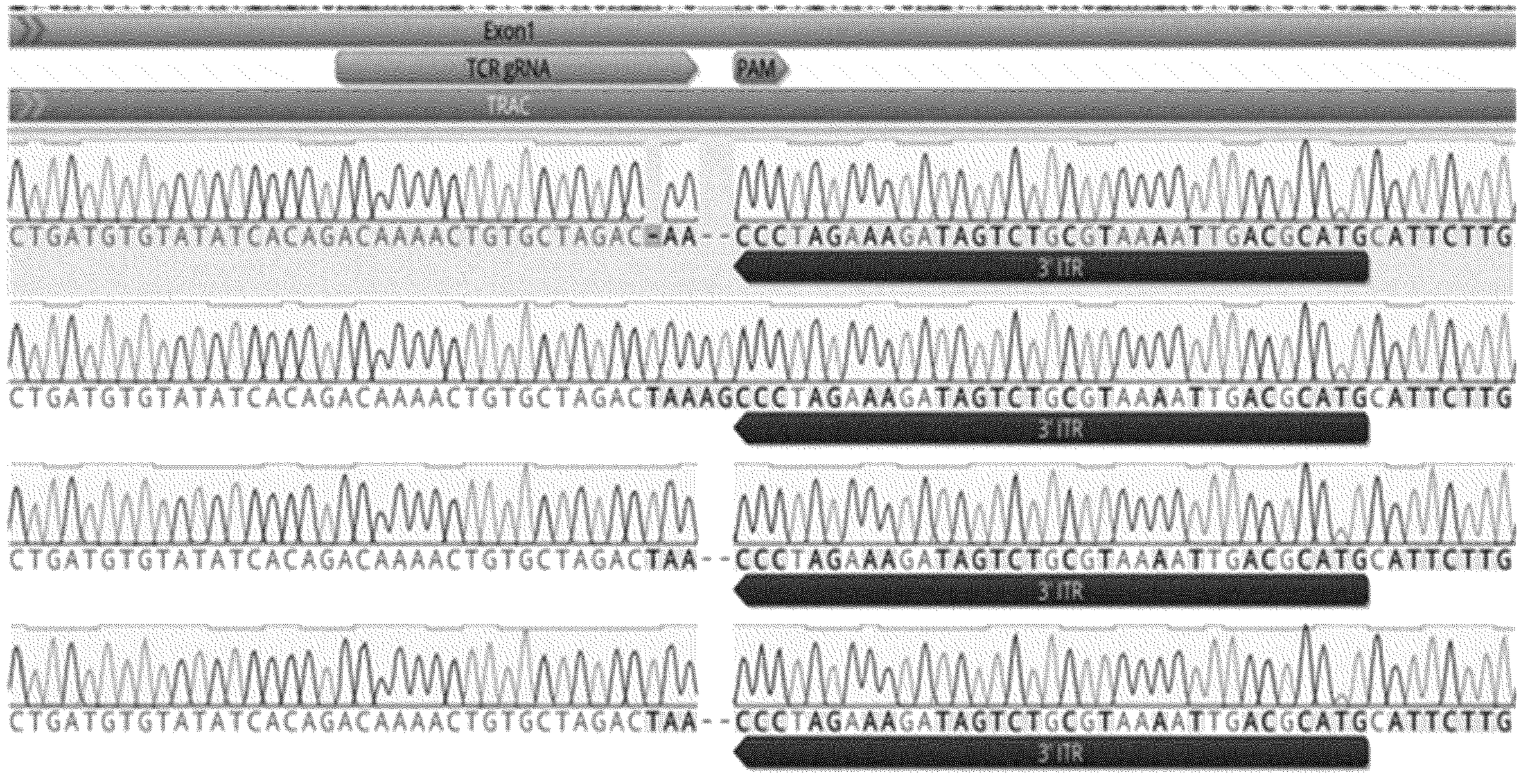
FIG. 8A

SUBSTITUTE SHEET (RULE 26)

Guide-seq profile of TCR-directed FiCAP* insertion (exon 1 TRAC loci):



Sanger sequencing of 4 clones:



* FiCAP: Cas9-hyPB R372A-K372A-D450N

FIG. 8B

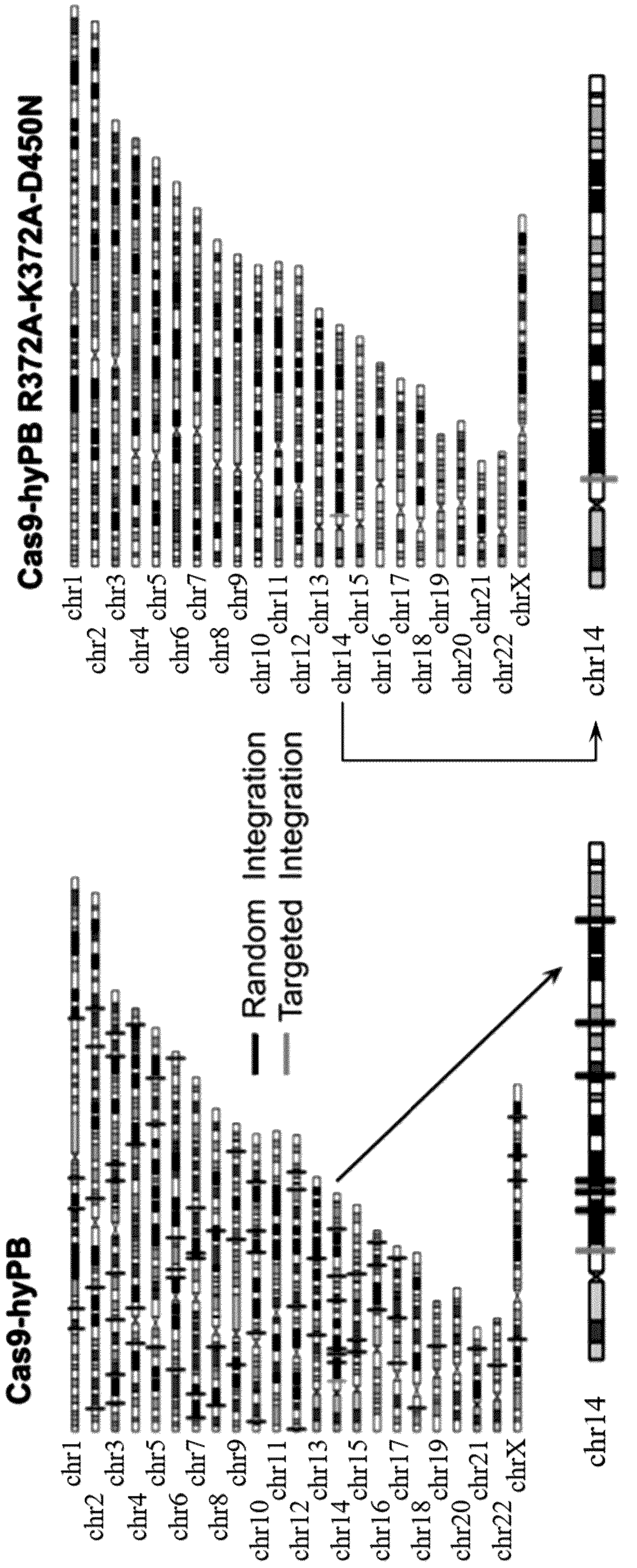

Cas9-hyPB R372A-K372A-D450N
Random Integration
Targeted Integration
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chr14
Cas9-hyPB
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chr14

FIG. 9

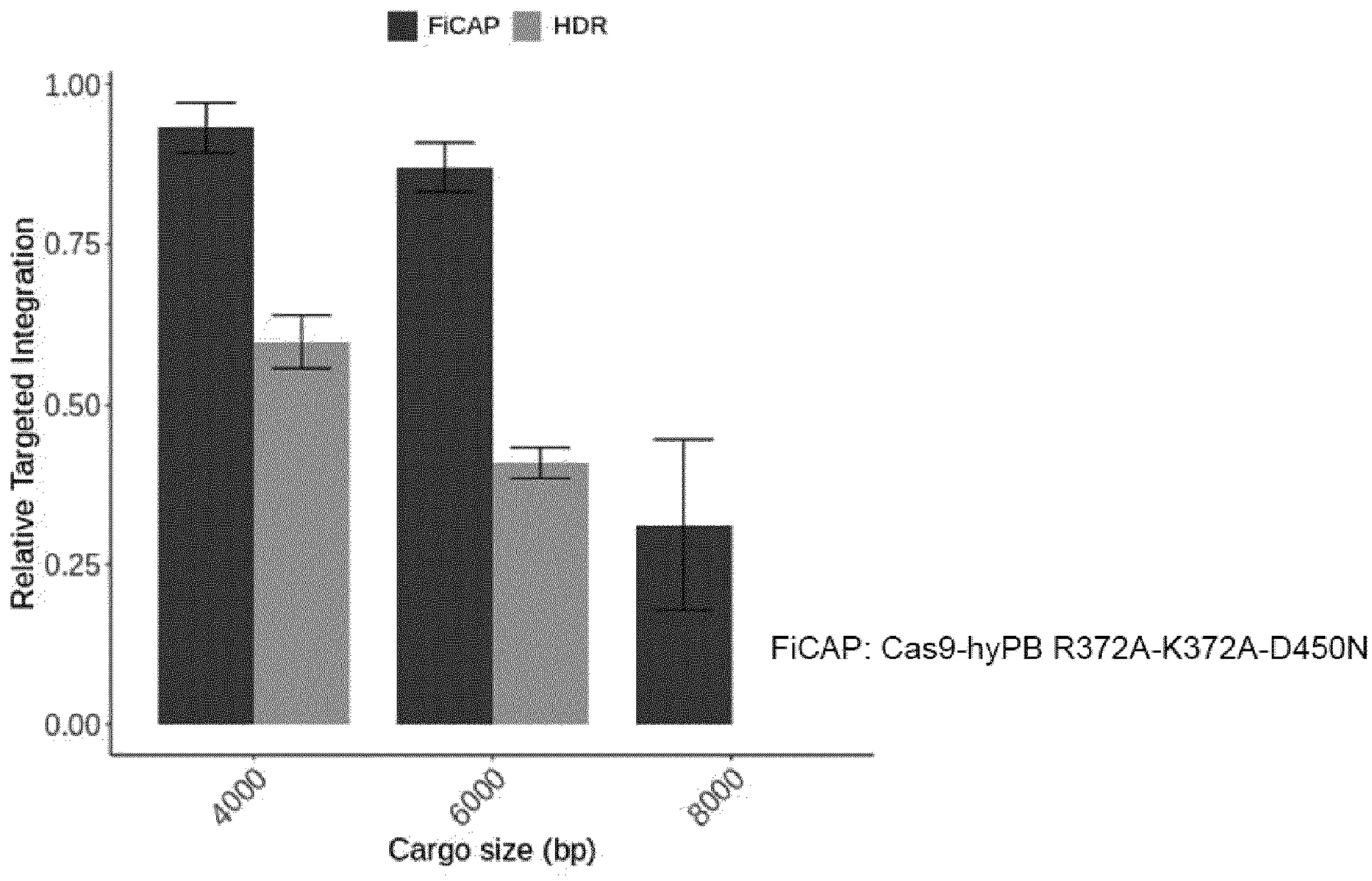

FiCAP
HDR
Relative Targeted Integration
1.00
0.75
0.50
0.25
0.00
4000
6000
8000
Cargo size (bp)
FiCAP: Cas9-hyPB R372A-K372A-D450N

FIG. 10

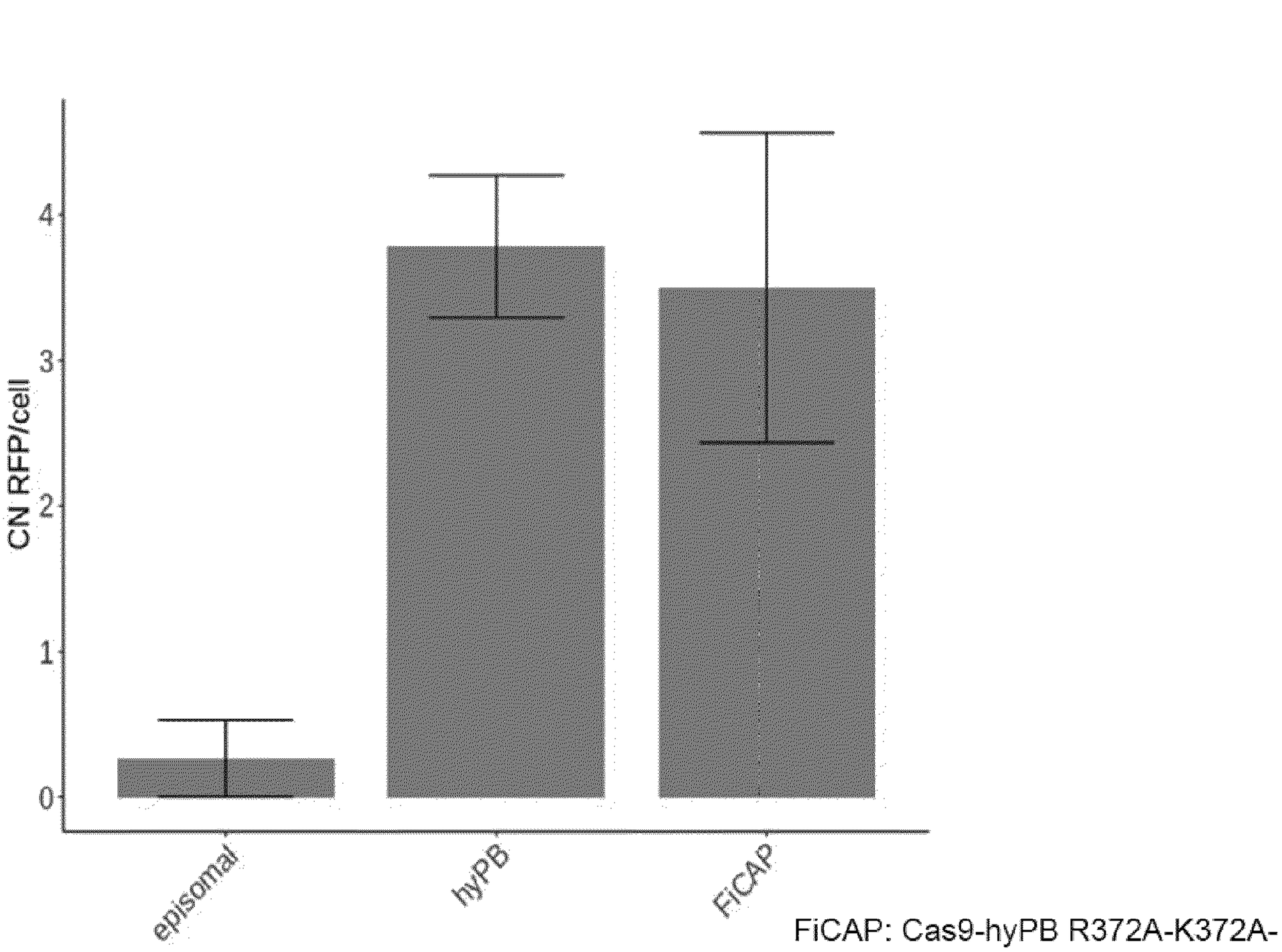

CN RFP/cell
4
3
2
1
0
episomal
hyPB
FiCAP
FiCAP: Cas9-hyPB R372A-K372A-

FIG. 11

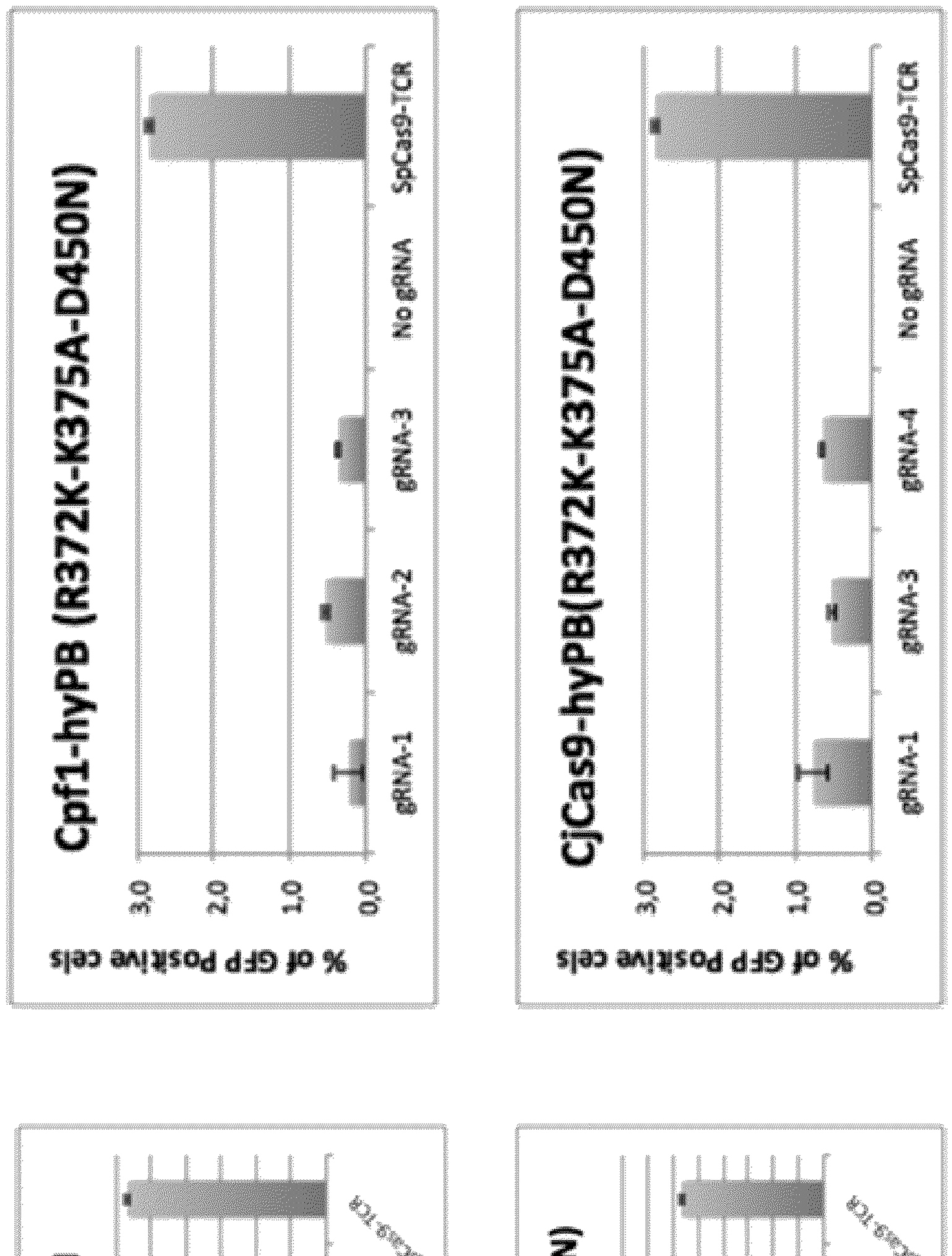
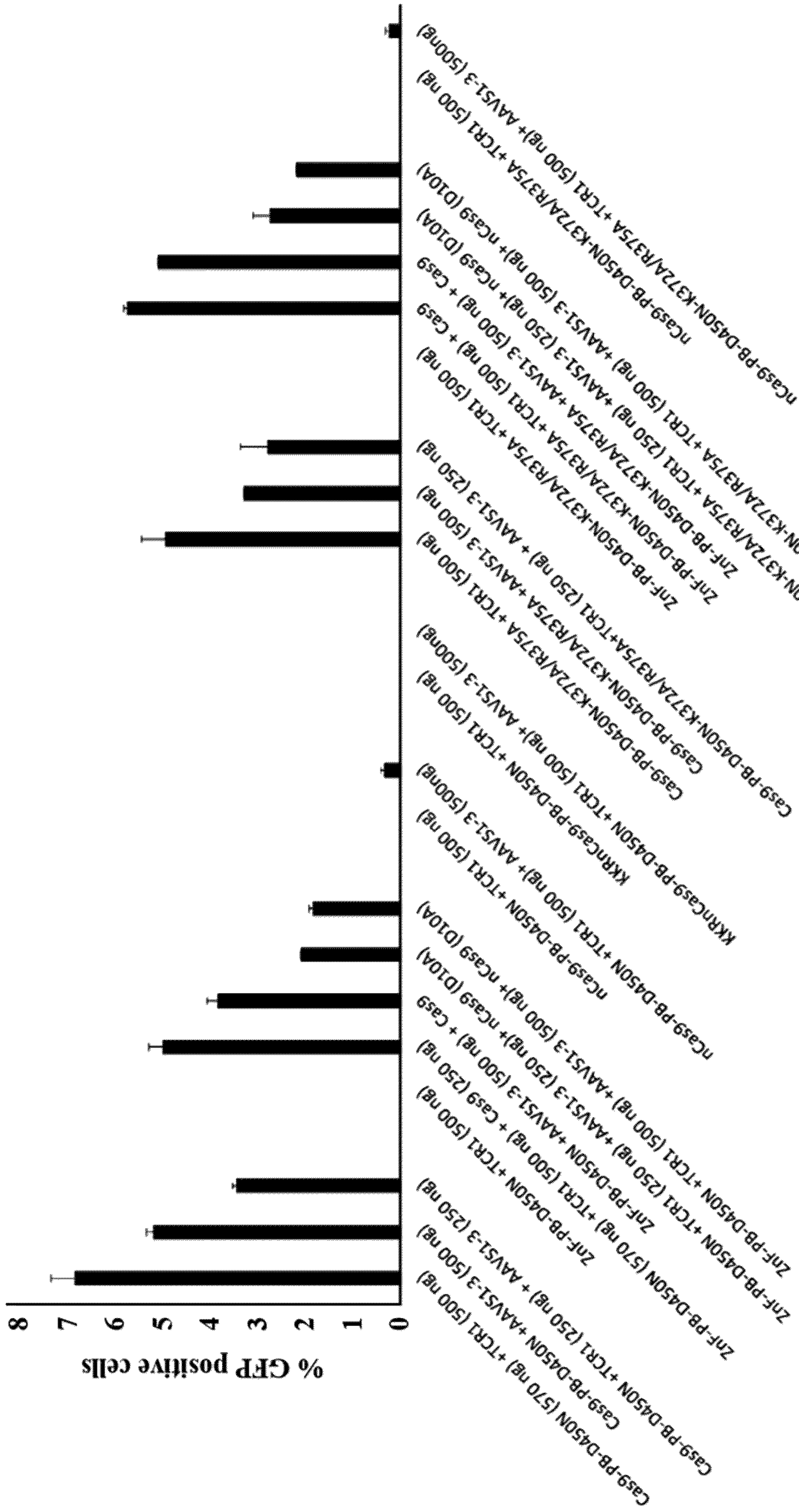

Cpf1-hyPB (R372K-K375A-D450N)
CjCas9-hyPB(R372K-K375A-D450N)
CasX-hyPB(R372A-K375A-D450N
SaCas9-hyPB (R372K-K375A-D450N)
% of GFP Positive cells
gRNA-1
gRNA-2
gRNA-3
gRNA-4
gRNA-5
No gRNA
SpCas9-TCR

FIG. 12


% GFP positive cells
0
1
2
3
4
5
6
7
8
Cas9-PB-D450N (570 ng) +TCR1 (500 ng)
Cas9-PB-D450N (570 ng) +AAV51-3 (250 ng)
Cas9-PB-D450N +TCR1 (250 ng) +AAV51-3 (500 ng)
ZnF-PB-D450N (570 ng) +TCR1 (500 ng)
ZnF-PB-D450N (570 ng) +AAV51-3 (250 ng)
ZnF-PB-D450N +TCR1 (250 ng) +AAV51-3 (500 ng)
nCas9-PB-D450N +TCR1 (500 ng)+ AAV51-3 (500ng)
KKRnCas9-PB-D450N +TCR1 (500 ng)+ AAV51-3 (500 ng)
Cas9-PB-D450N +TCR1 (500 ng)+ AAV51-3 (500 ng)
nCas9-PB-D450N +TCR1 (500 ng)+ AAV51-3 (500 ng)
ZnF-PB-D450N +TCR1 (250 ng) +AAV51-3 (500 ng)
ZnF-PB-D450N (570 ng) +TCR1 (500 ng)
ZnF-PB-D450N (570 ng) +AAV51-3 (250 ng)
Cas9-PB-D450N +TCR1 (250 ng) +AAV51-3 (500 ng)
nCas9-PB-D450N +TCR1 (500 ng)+ AAV51-3 (500

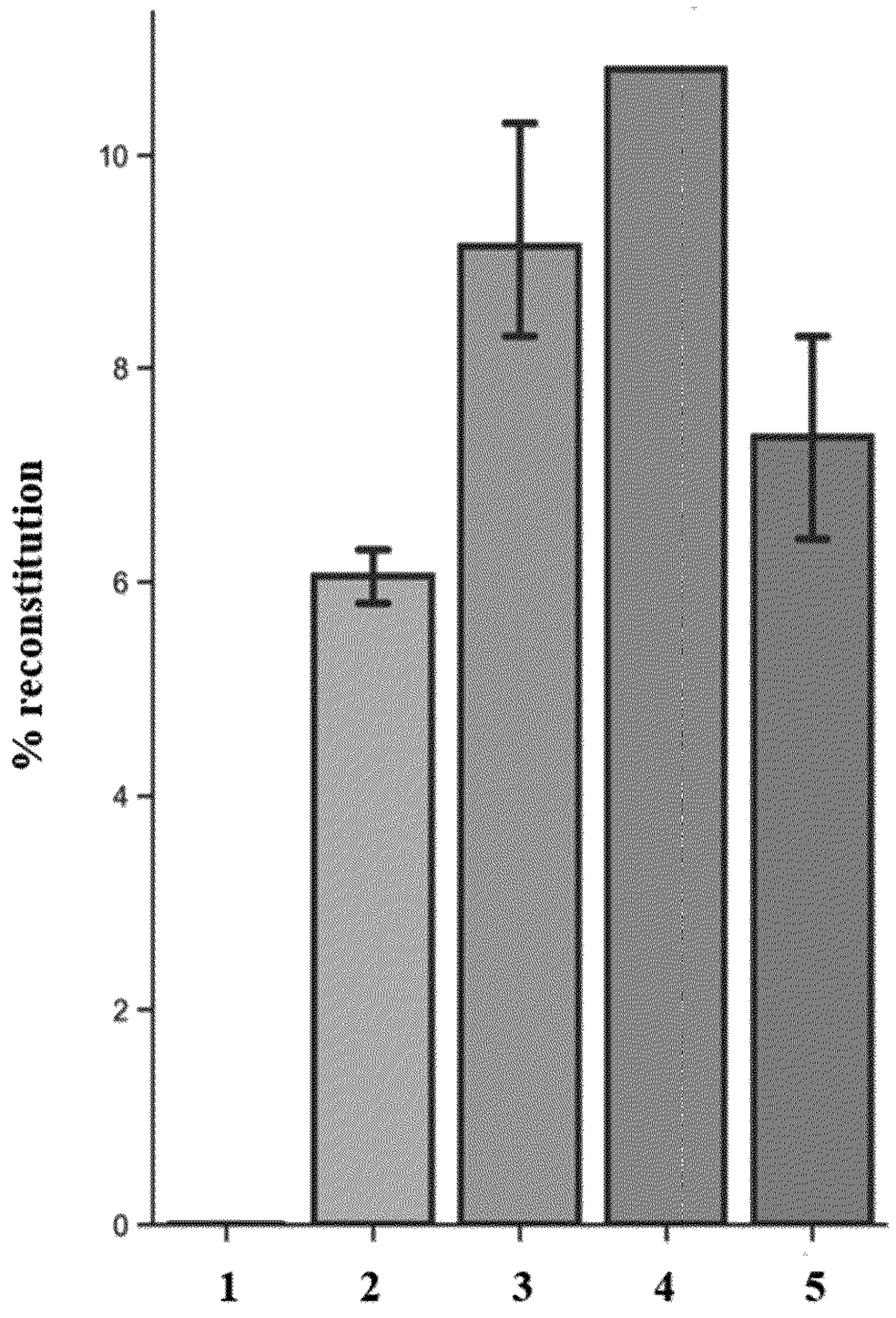


FIG. 14

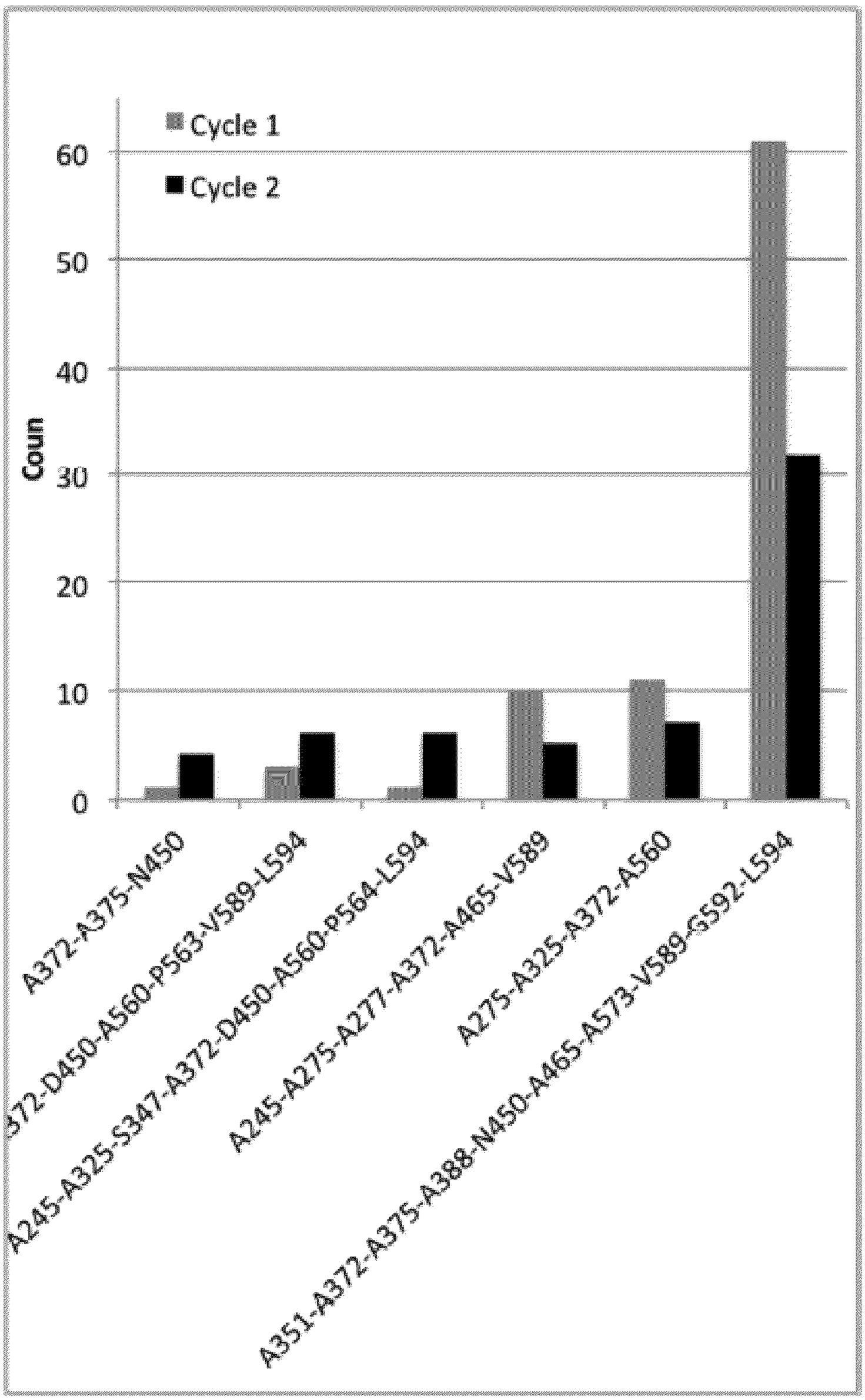

Cycle 1
Cycle 2
60
50
40
30
20
10
0
Coun
A372-A375-N450
A372-D450-A560-P563-V589-L594
A245-A325-S347-A372-D450-A560-P564-L594
A245-A275-A277-A372-A465-V589
A275-A325-A372-A560
A351-A372-A375-A388-N450-A465-A573-V589-G592-L594

FIG. 15

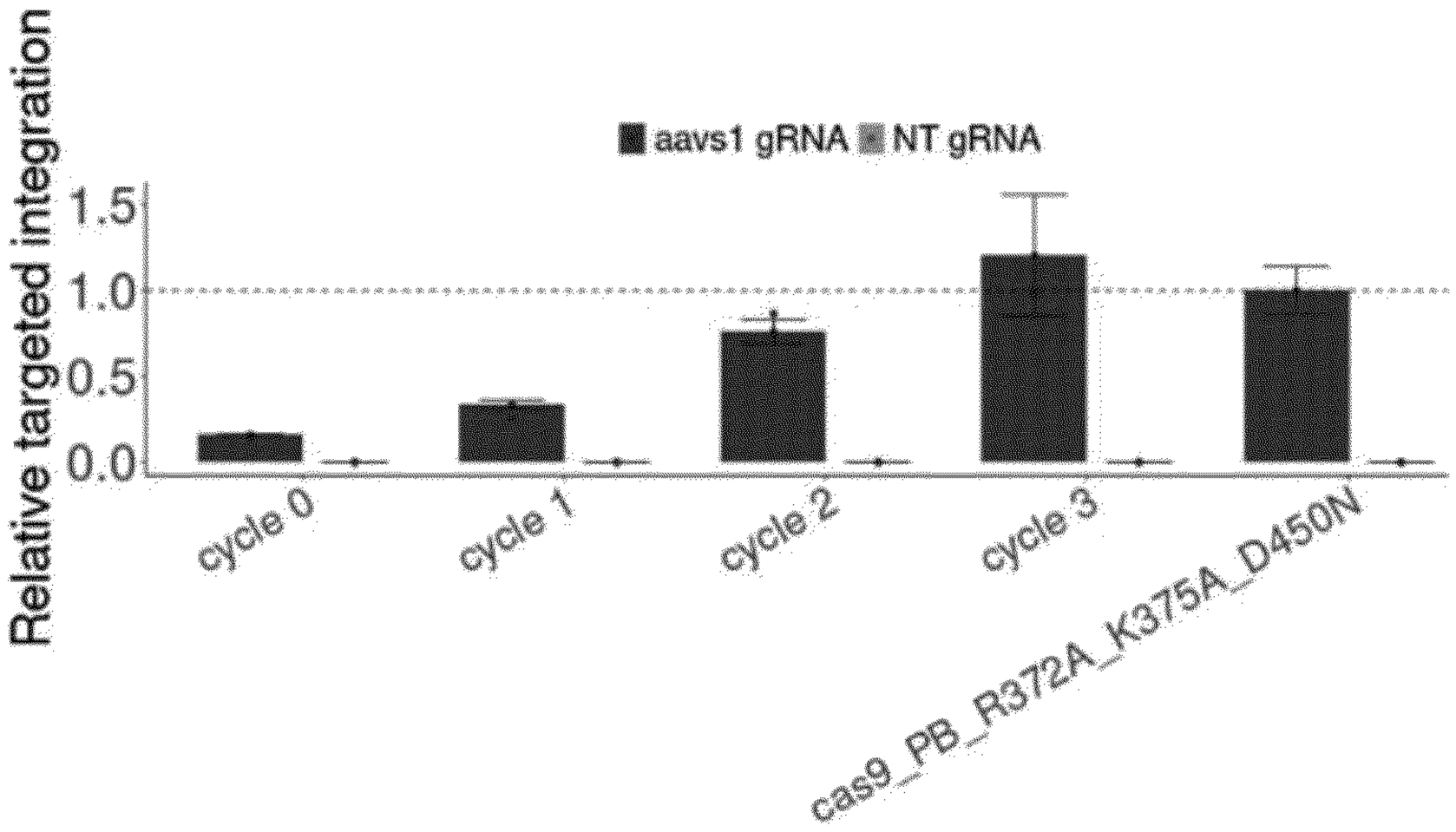


FIG. 16

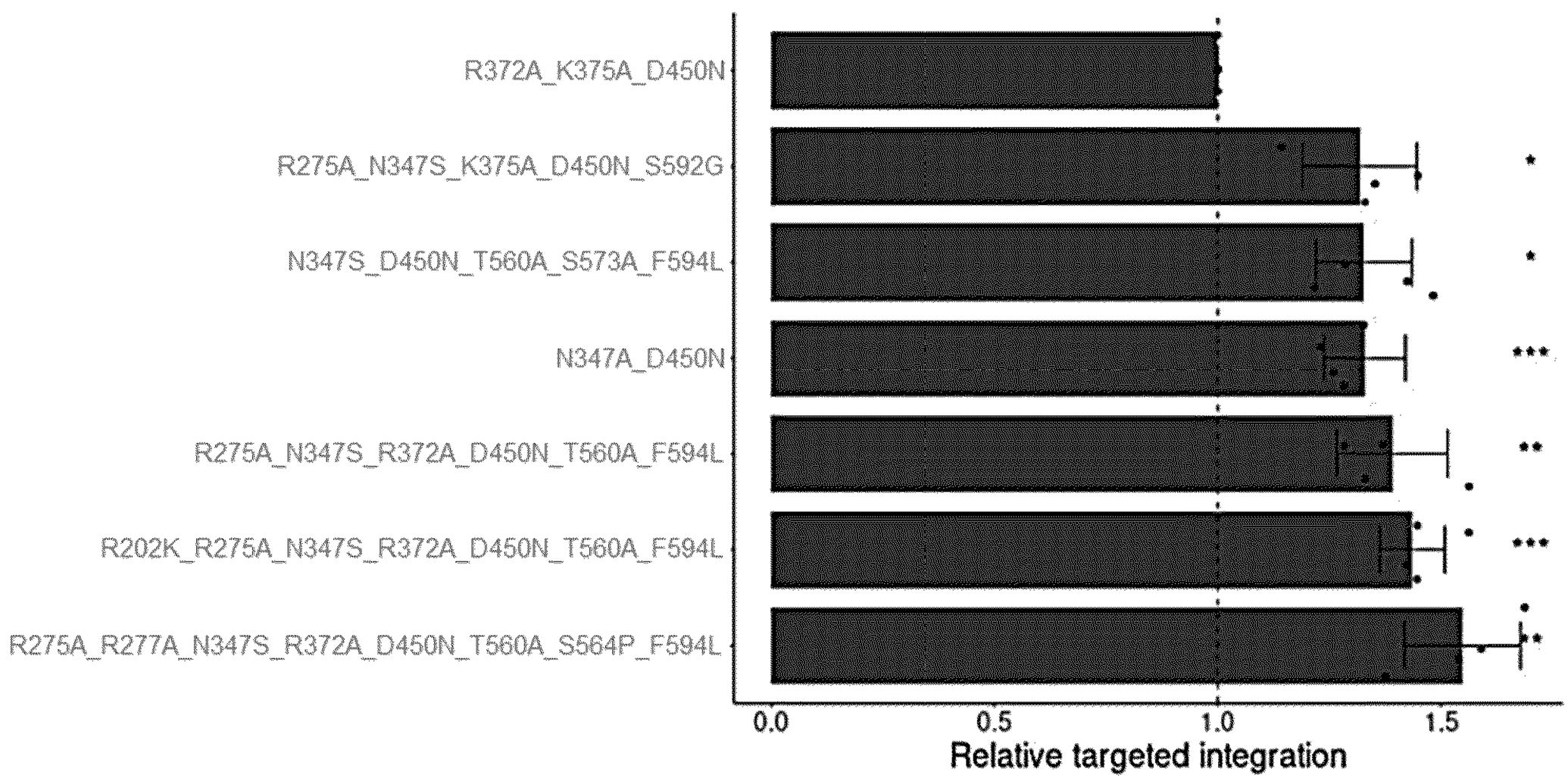


FIG. 17A



FIG. 17B

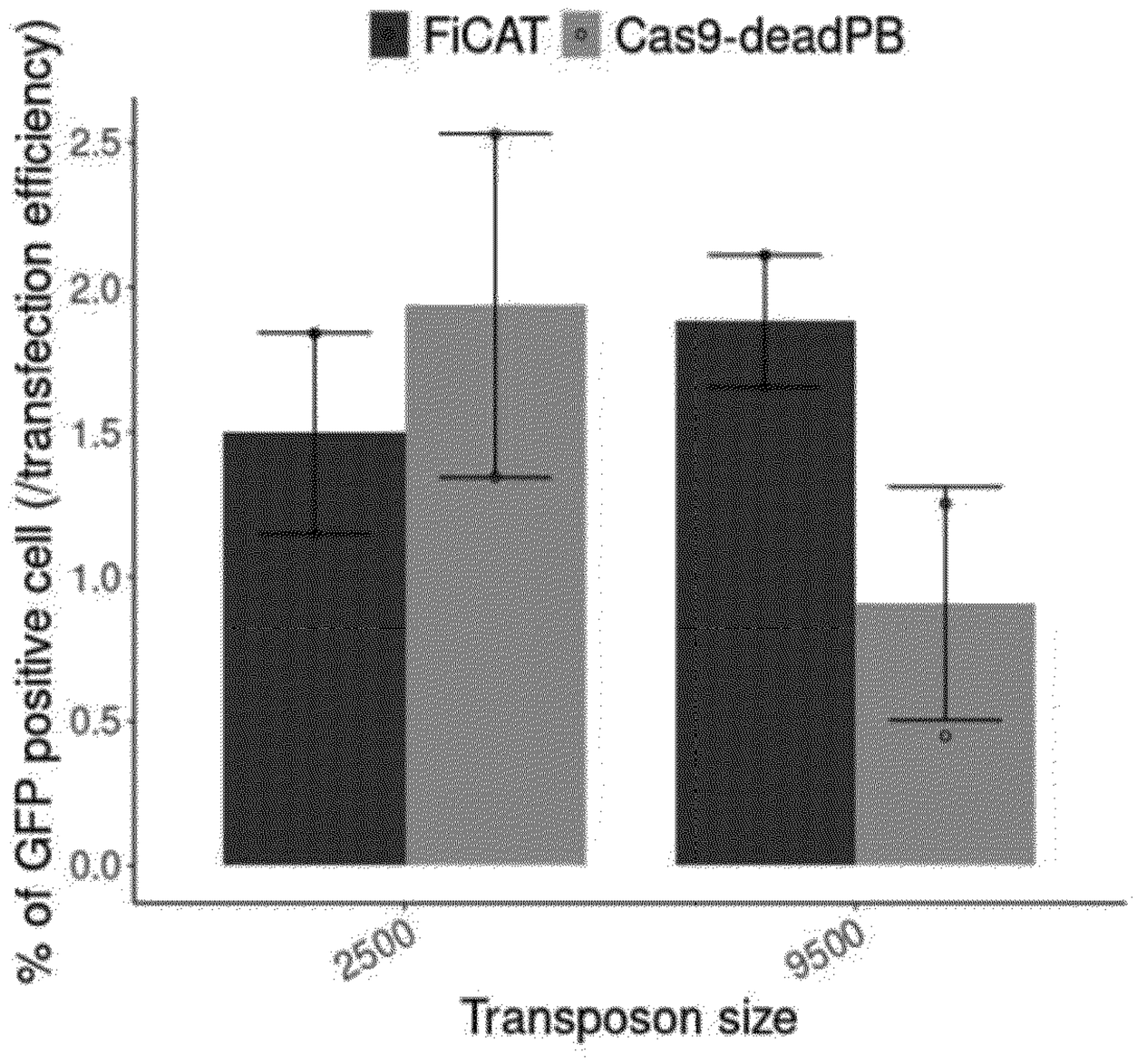


FIG. 18

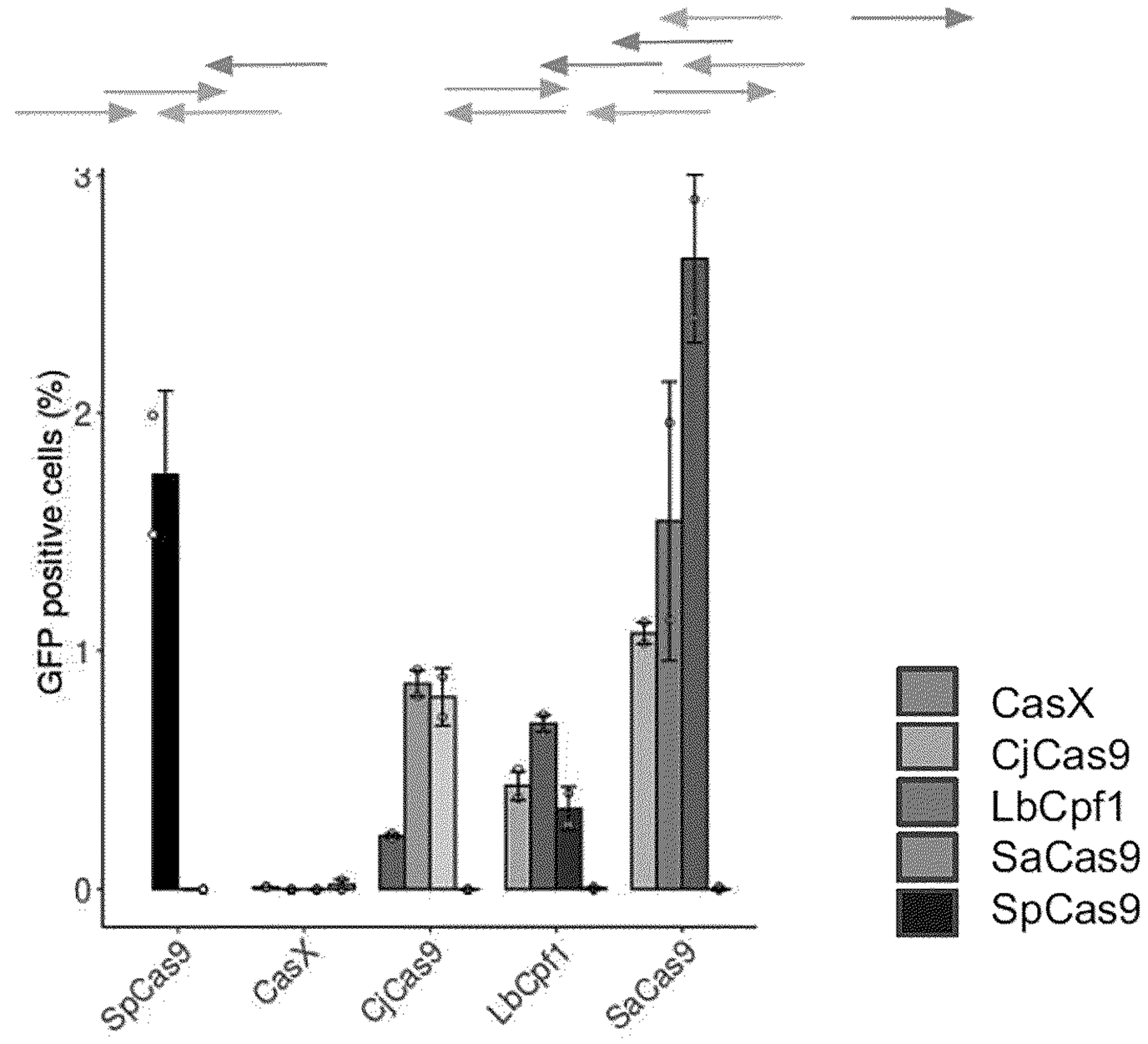

GENOME INTEGRATED REPORTER
GFP positive cells (%)
0
1
2
3
SpCas9
CasX
CjCas9
LbCpf1
SaCas9
CasX
CjCas9
LbCpf1
SaCas9
SpCas9

FIG. 19

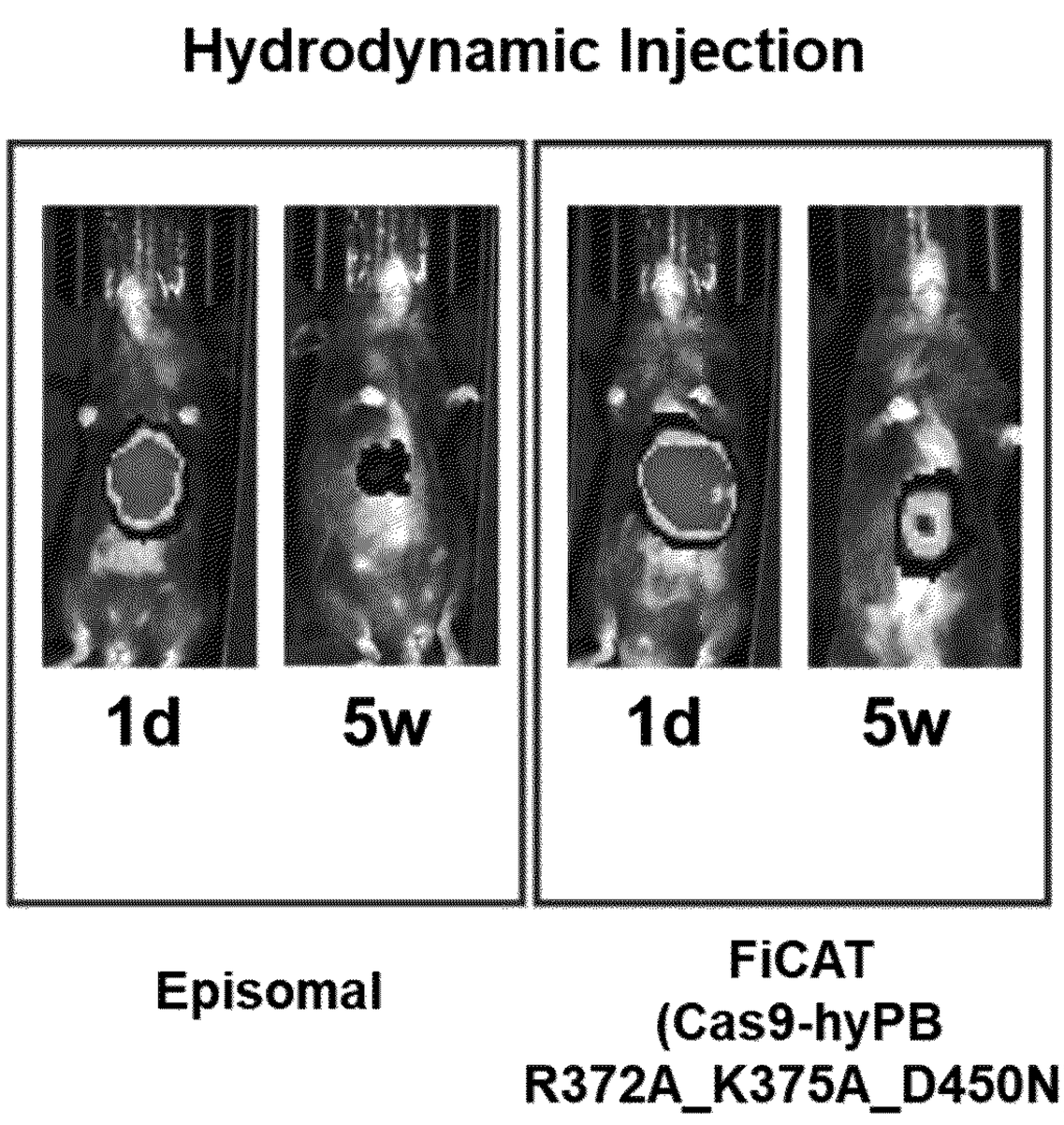

Hydrodynamic Injection
1d
5w
1d
5w
Episomal
FiCAT
(Cas9-hyPB
R372A_K375A_D450N)

FIG. 20

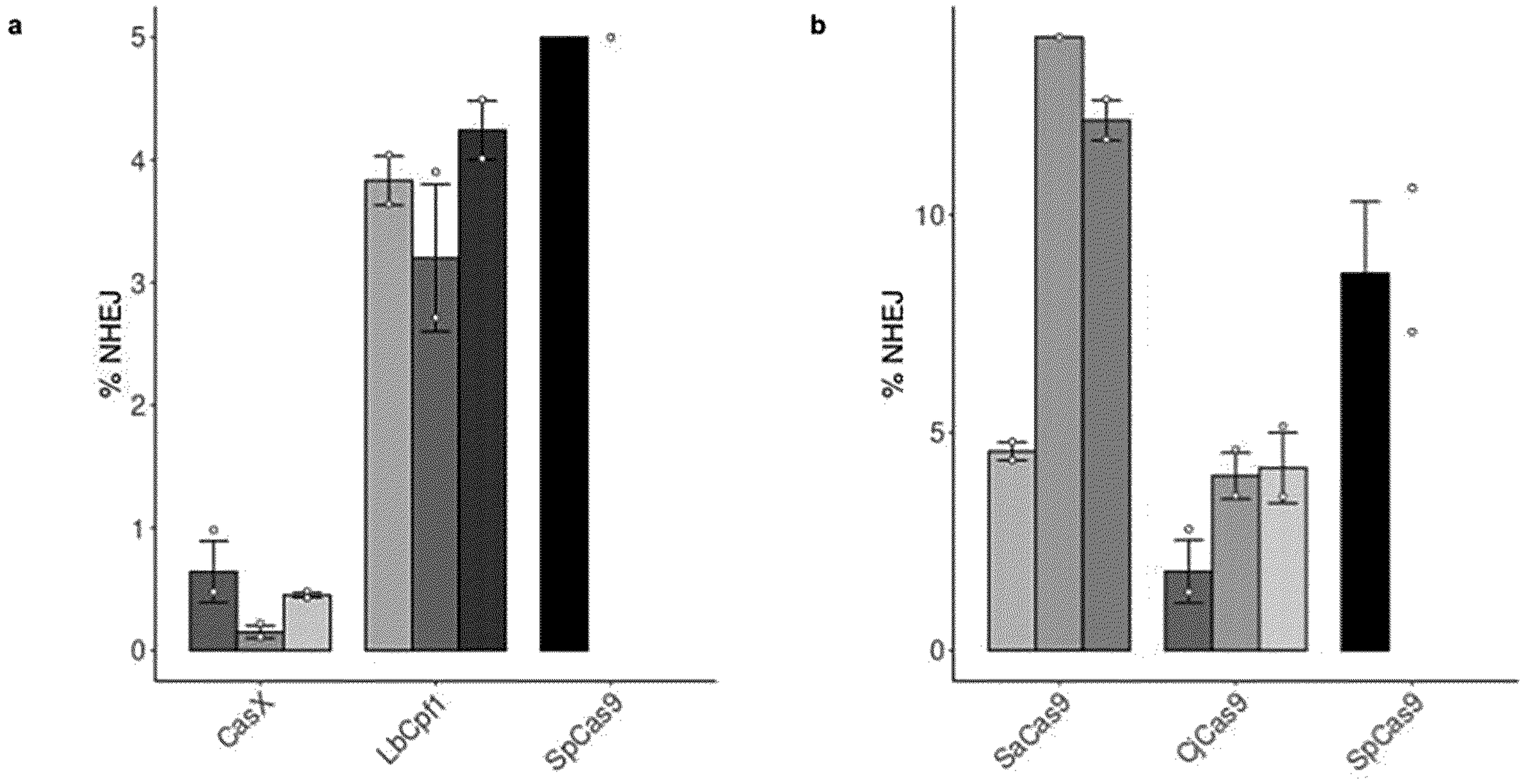


FIG. 21

12

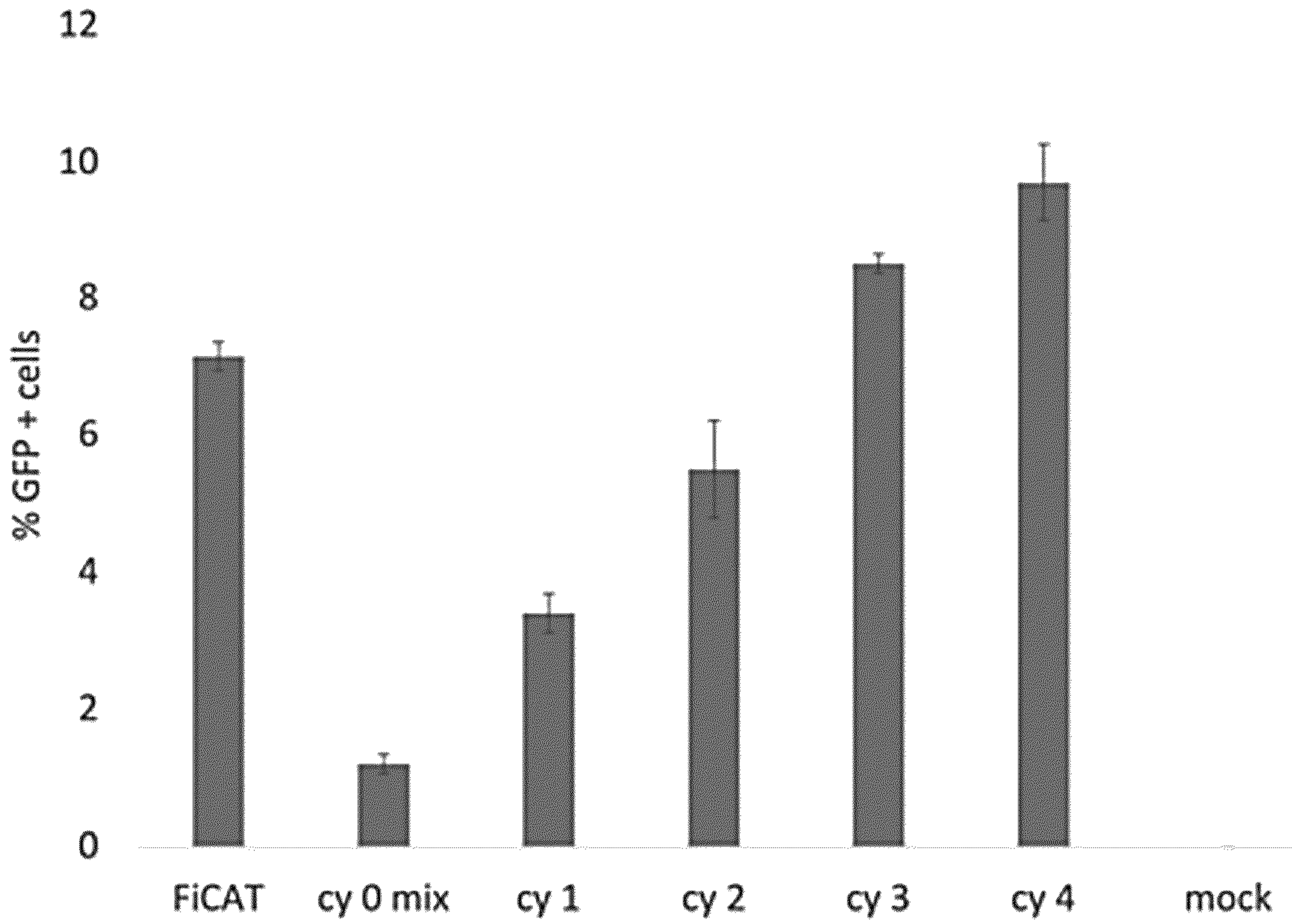


FIG. 22A

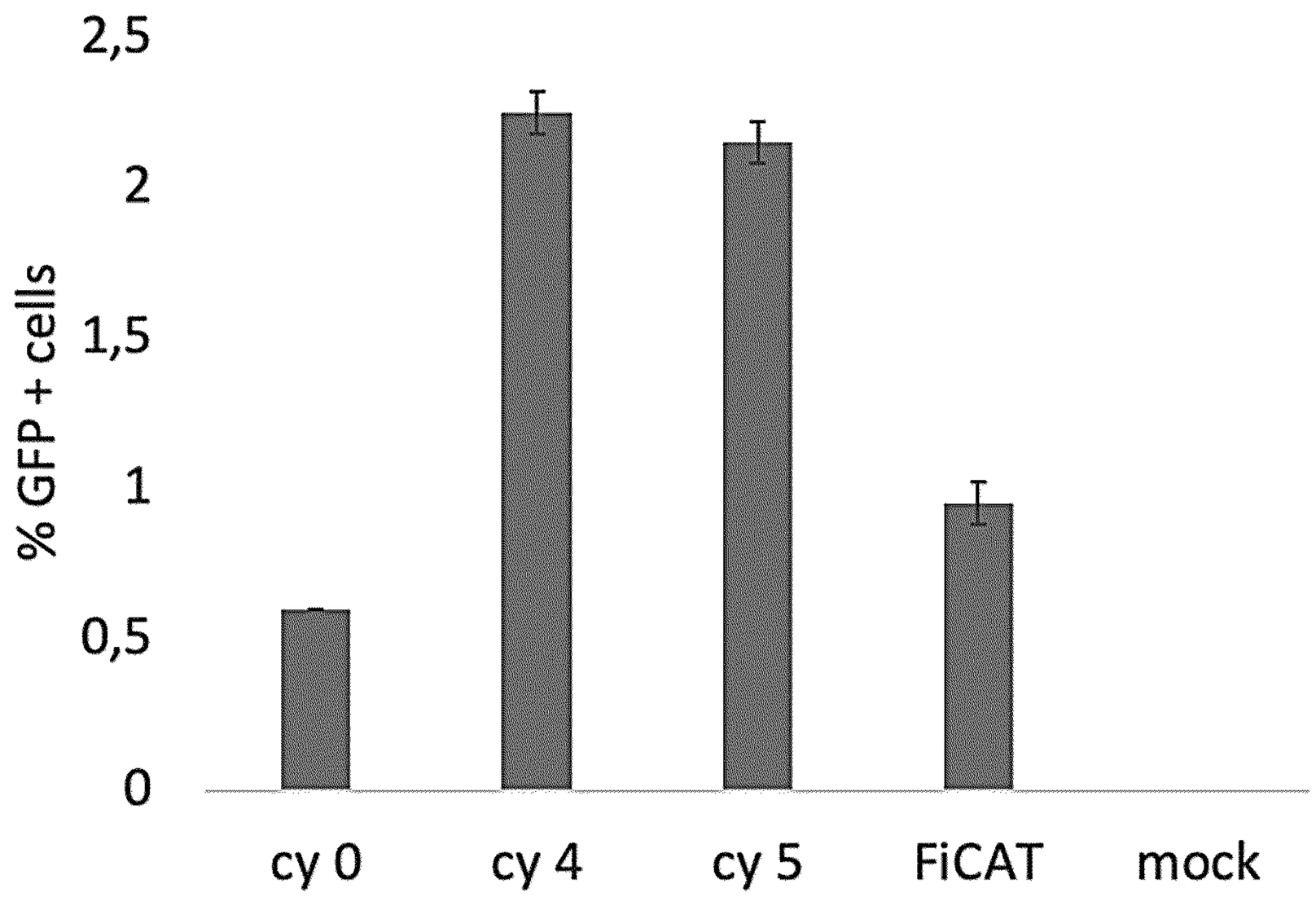


FIG. 22B

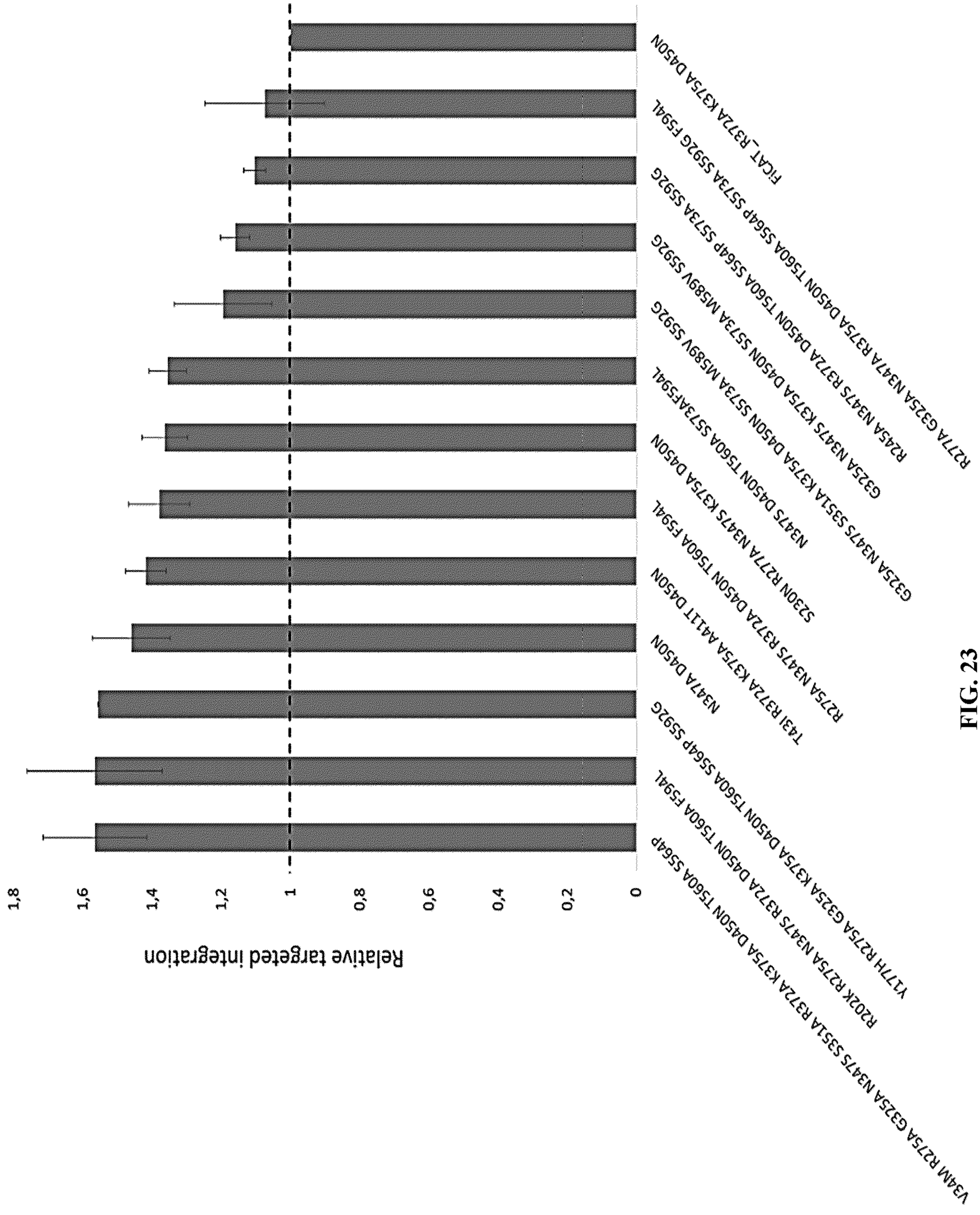

Relative targeted integration
0
0,2
0,4
0,6
0,8
1
1,2
1,4
1,6
1,8
FICAT_R372A K375A D45ON
G325A N347S S351A K375A D45ON T560A F594L
R245A N347S R372A D45ON T560A S564P S573A M589V S592G
R277A G325A N347A R375A D45ON T560A S564P S573A S592G
G325A N347S S351A R372A D45ON T560A S564P
R202K R275A N347S R372A D45ON T560A F594L
V177H R275A G325A K375A D45ON T560A S564P
V34M R275A G325A N347S S351A R372A K375A D45ON T560A S564P
R275A N347S R372A A411T D45ON
S230N R277A N347S K375A D45ON T560A F594L
N347S D45ON T560A K375A D45ON
G325A N347S S351A K375A D45ON S573A M589V S592G
R245A N347S R372A D45ON S573A M589V S592G
T431 R372A K375A A411T D45ON
N347A D45ON
S564P S592G

FIG. 23

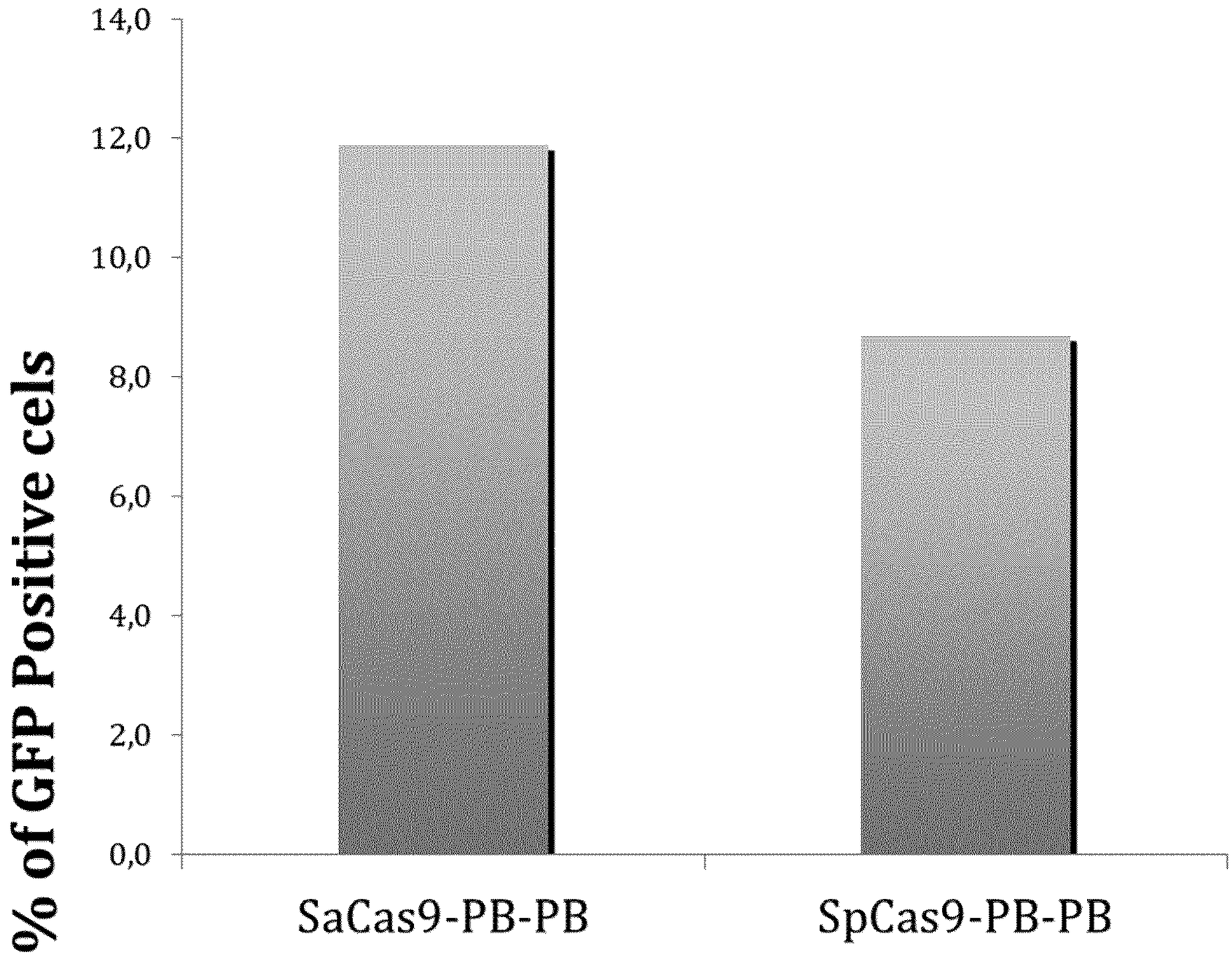


FIG. 24

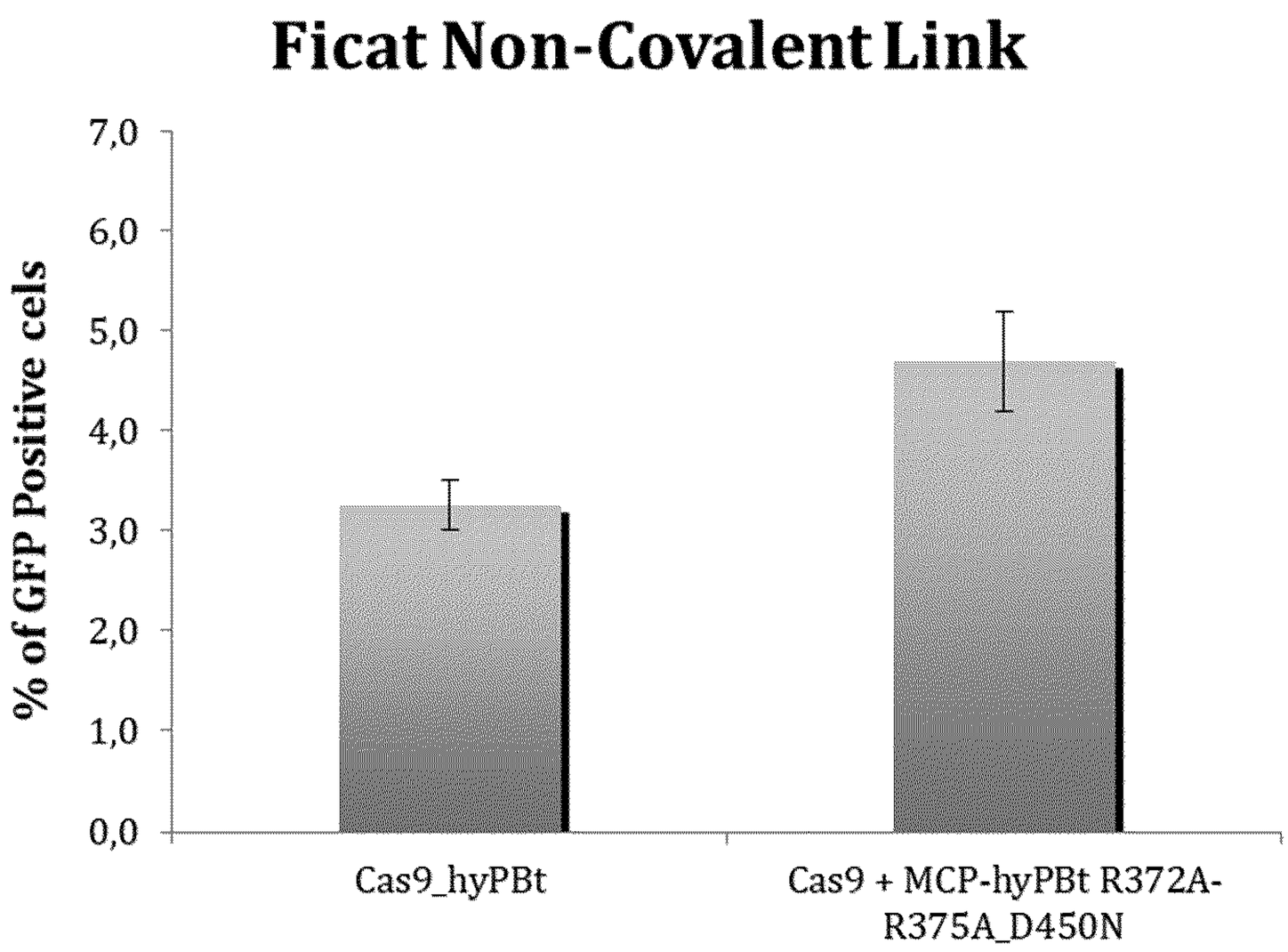


FIG. 25A

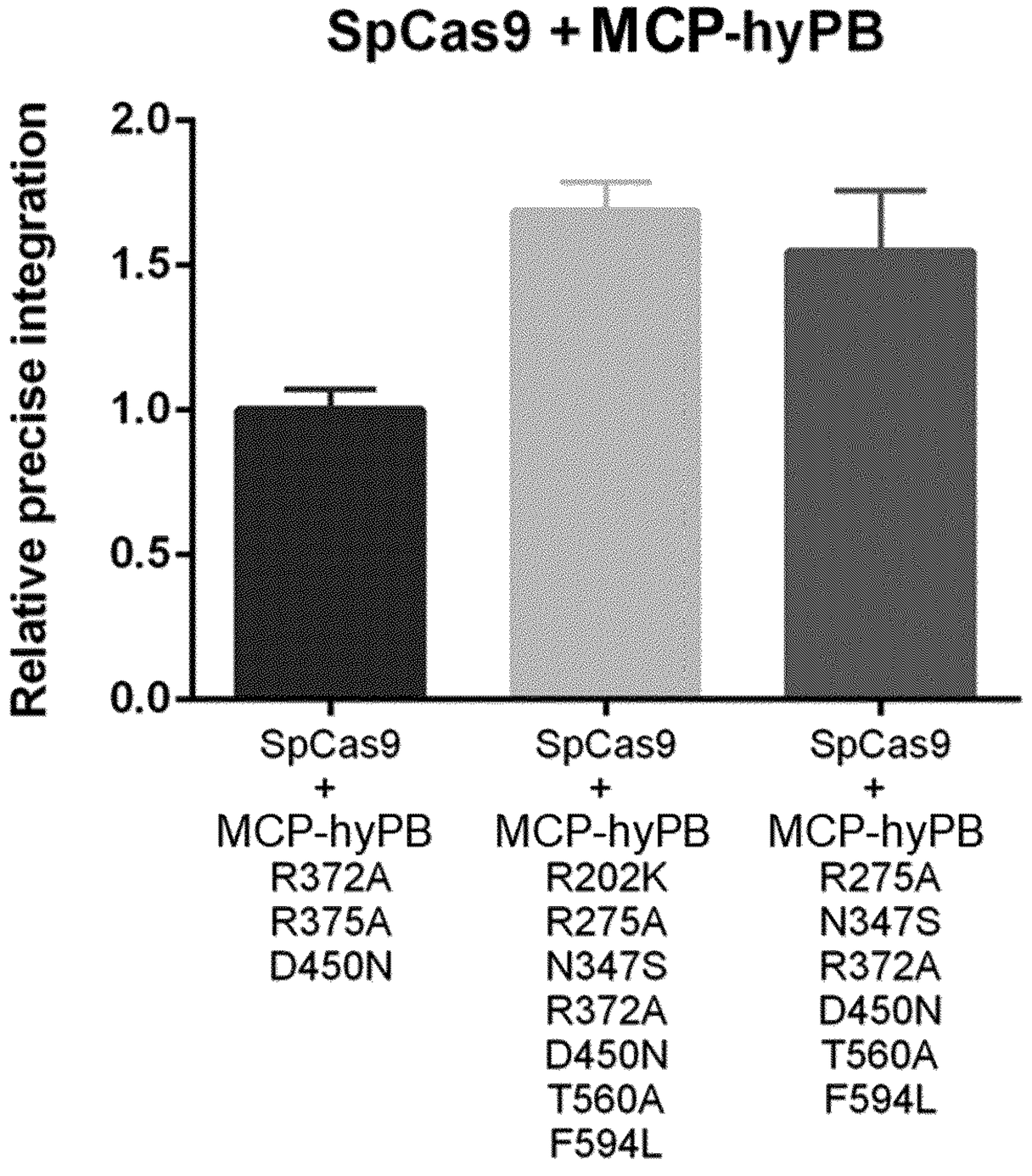


FIG. 25B

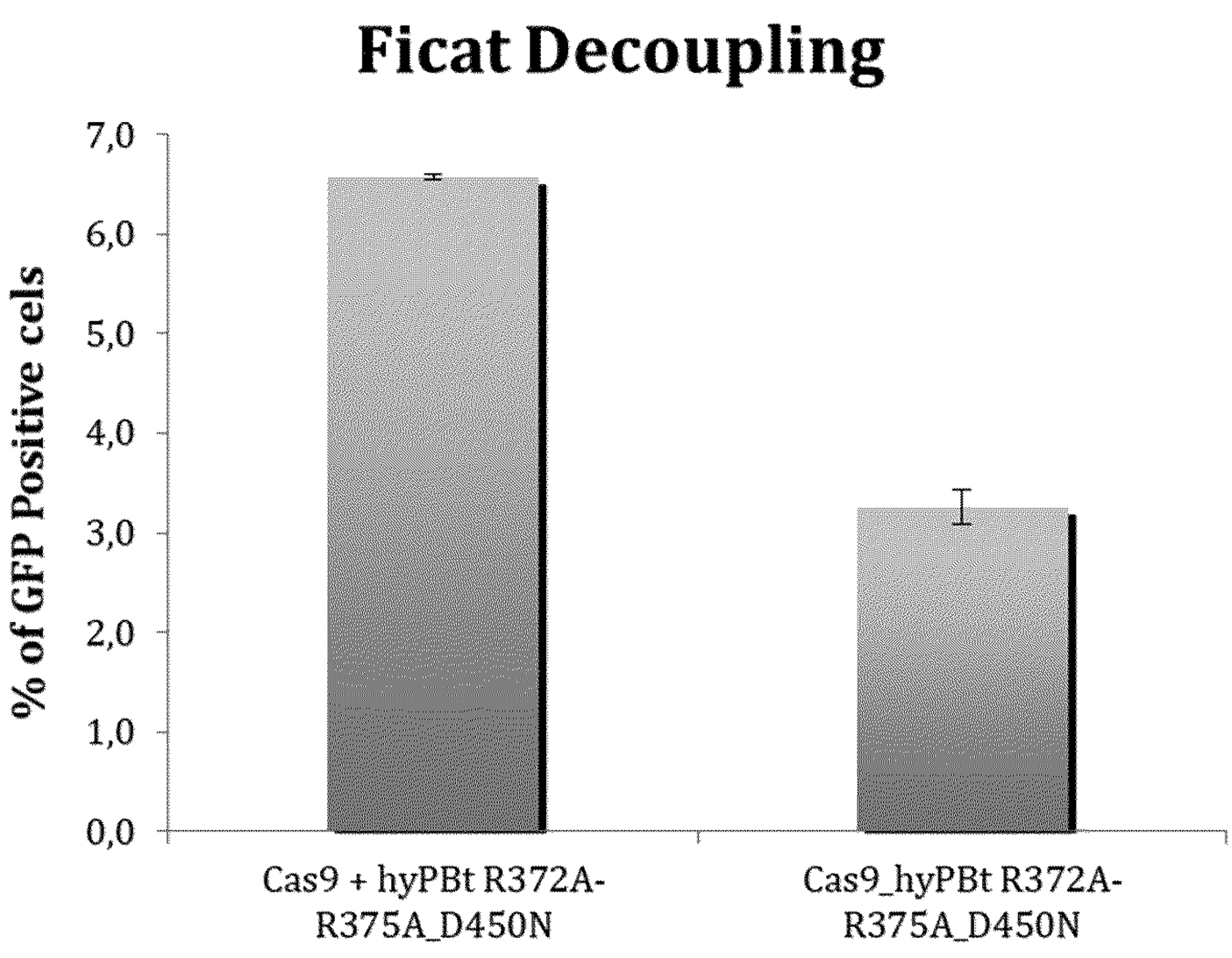


FIG. 26A

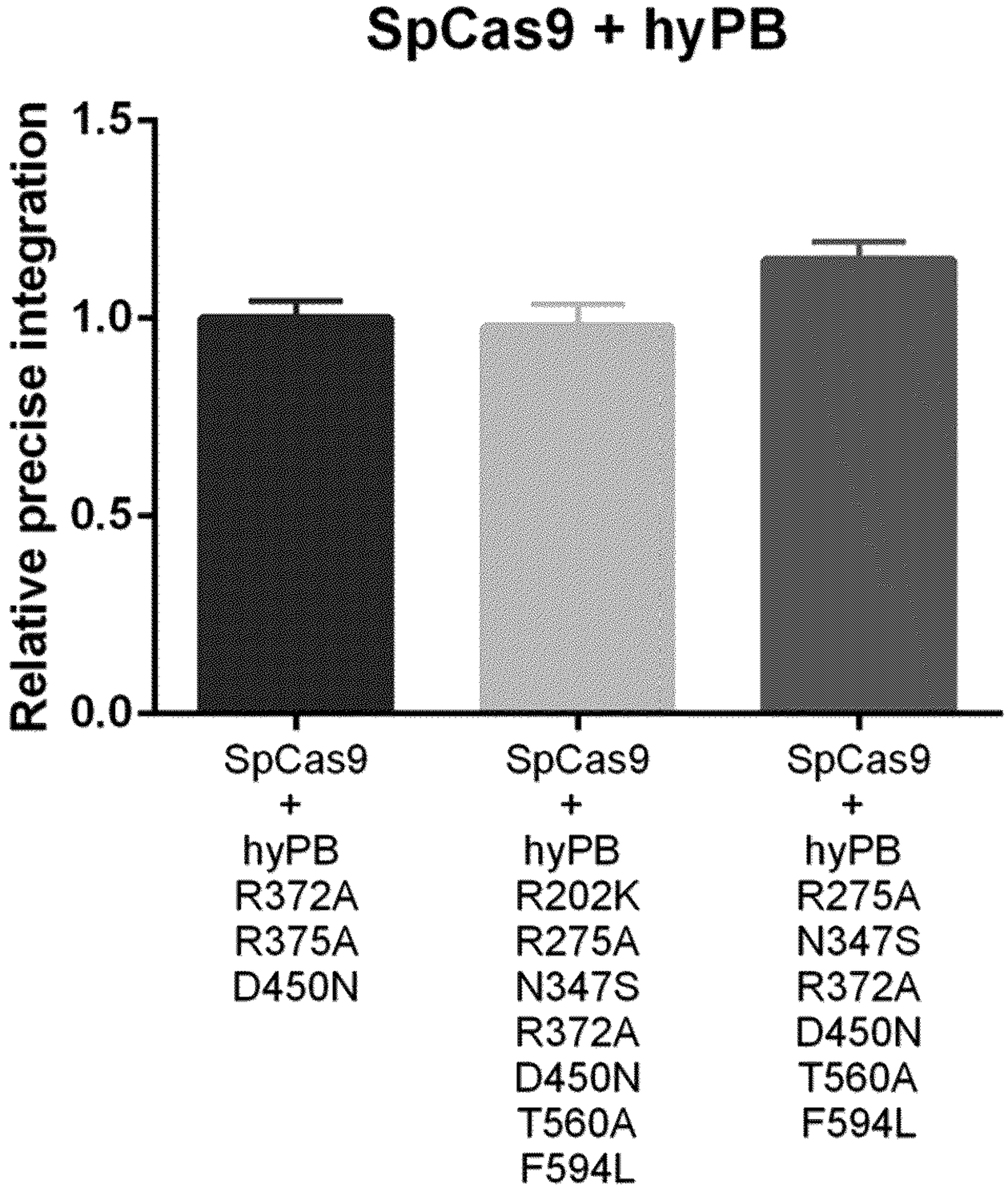

SpCas9 + hyPB
Relative precise integration
1.5
1.0
0.5
0.0
SpCas9 + hyPB R372A R375A D450N
SpCas9 + hyPB R202K R275A N347S R372A D450N T560A F594L
SpCas9 + hyPB R275A N347S R372A D450N T560A F594L

FIG. 26B

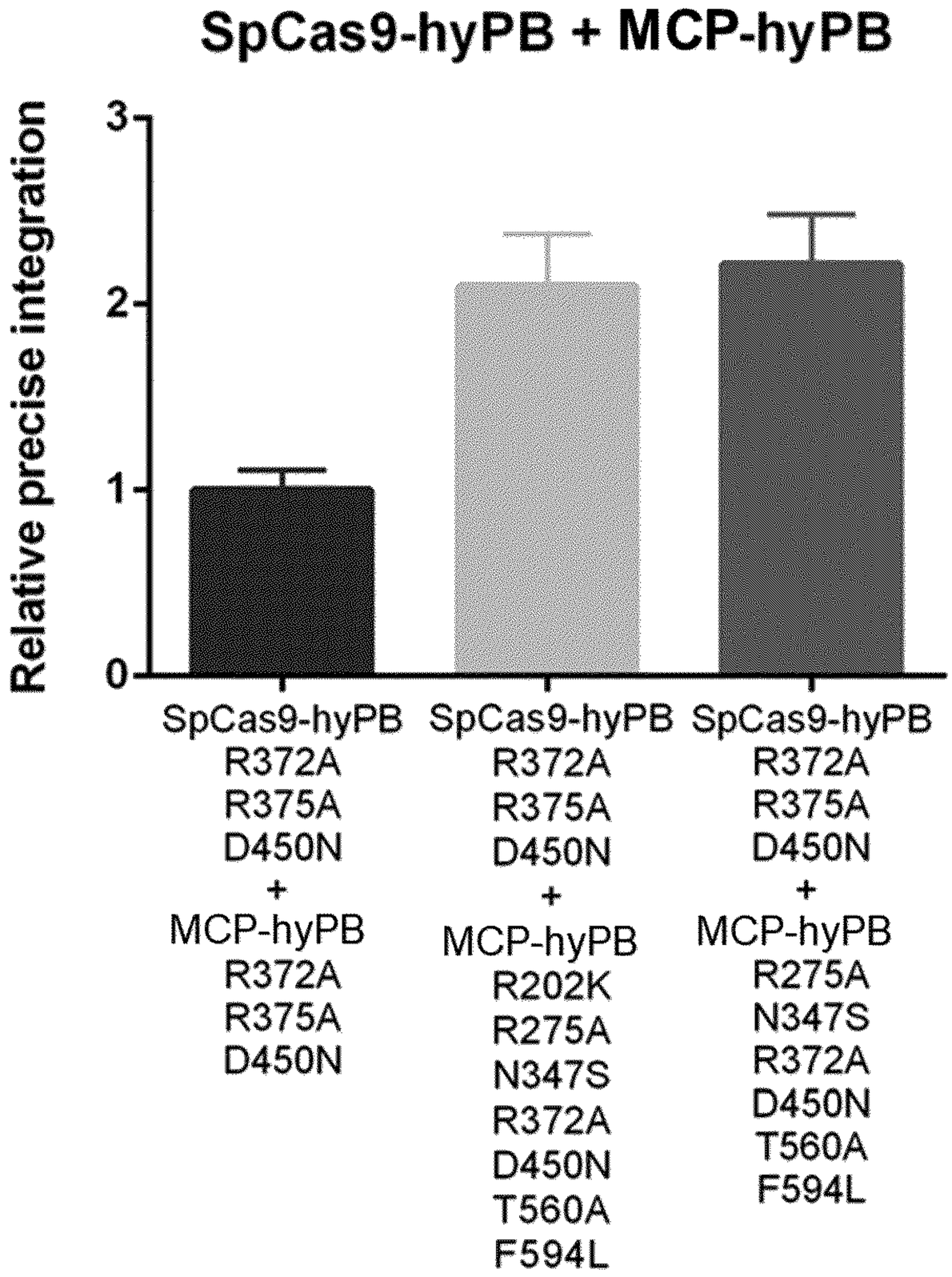

SpCas9-hyPB + MCP-hyPB
Relative precise integration
3
2
1
0
SpCas9-hyPB R372A R375A D450N + MCP-hyPB R372A R375A D450N
SpCas9-hyPB R372A R375A D450N + MCP-hyPB R202K R275A N347S R372A D450N T560A F594L
SpCas9-hyPB R372A R375A D450N + MCP-hyPB R275A N347S R372A D450N T560A F594L

FIG 27

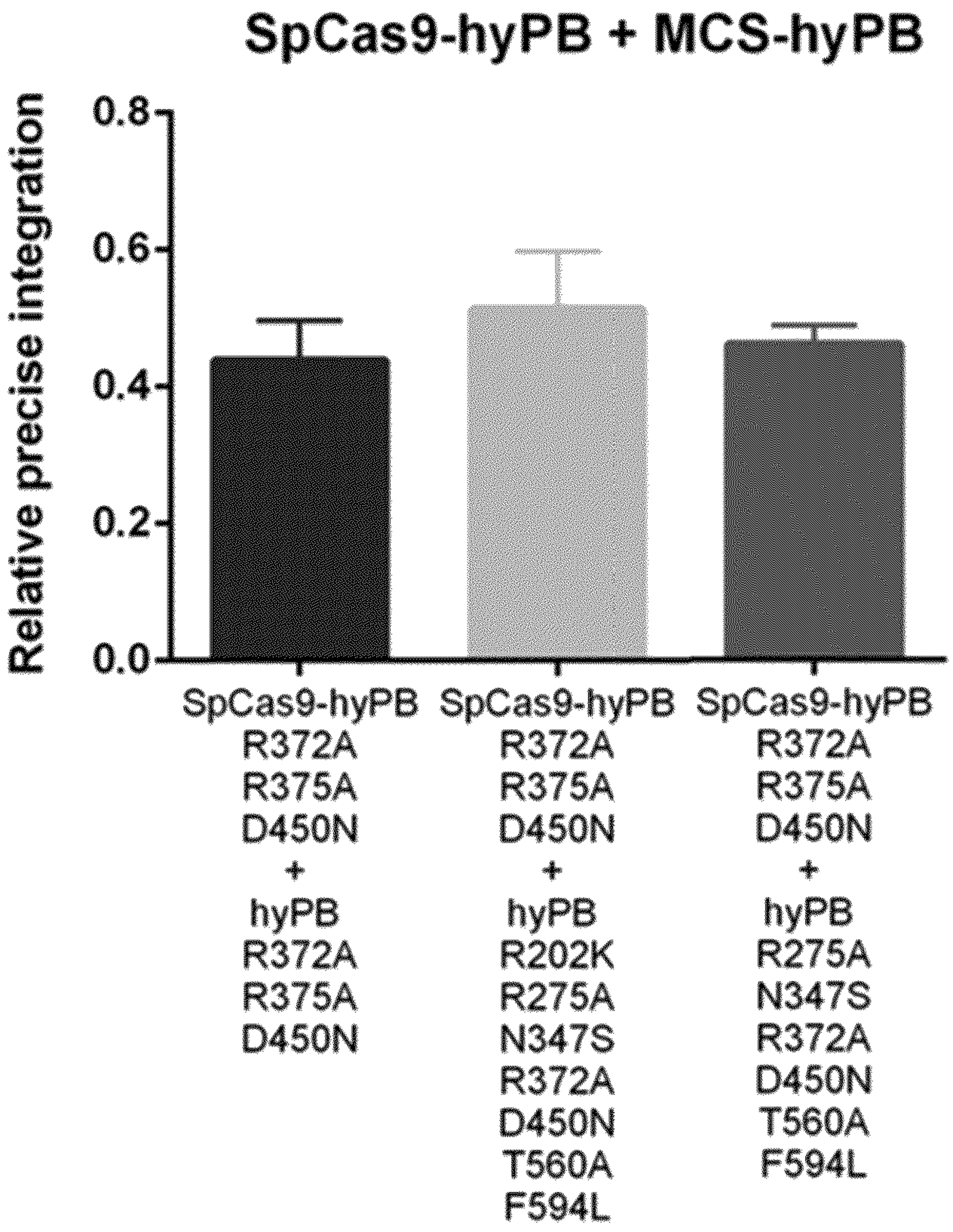

SpCas9-hyPB + MCS-hyPB
Relative precise integration
0.8
0.6
0.4
0.2
0.0
SpCas9-hyPB R372A R375A D450N + hyPB R372A R375A D450N R372A D450N T560A F594L
SpCas9-hyPB R372A R375A D450N + hyPB R202K R275A N347S D450N T560A F594L
SpCas9-hyPB R372A R375A D450N + hyPB R275A N347S R372A D450N T560A F594L

FIG. 28

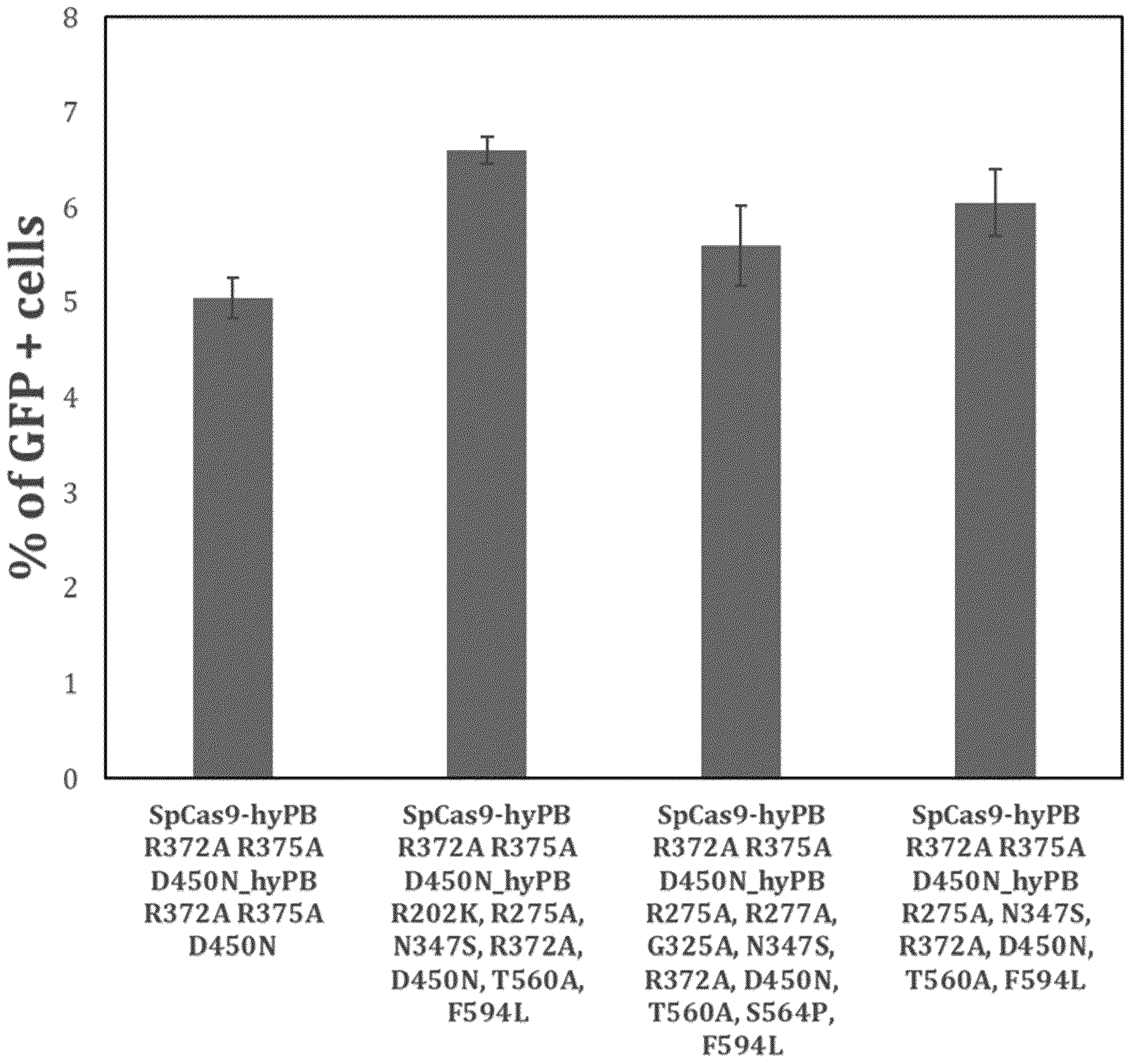

% of GFP + cells
0
1
2
3
4
5
6
7
8
SpCas9-hyPB R372A R375A D450N_hyPB R372A R375A D450N
SpCas9-hyPB R372A R375A D450N_hyPB R202K, R275A, N347S, R372A, D450N, T560A, F594L
SpCas9-hyPB R372A R375A D450N_hyPB R275A, R277A, G325A, N347S, R372A, D450N, T560A, S564P, F594L
SpCas9-hyPB R372A R375A D450N_hyPB R275A, N347S, R372A, D450N, T560A, F594L

FIG. 29

SEQUENCE LISTING

<110> Universitat Pompeu Fabra

<120> PROGRAMMABLE TRANSPOSASES AND USES THEREOF

<130> IBIO-2199/PCT

<150> EP21209719.0
<151> 2021-11-22
<150> EP20214696.5
<151> 2020-12-16

<160> 153

<170> BiSSAP 1.3.6

<210> 1
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified Hyperactive Piggybac

<400> 1
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn

```
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Ala Ser Asn Ala Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asn Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
```

580 585 590
Cys Phe

<210> 2
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<400> 2
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
245 250 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
260 265 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
275 280 285

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

<210> 3
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified Hyperactive Piggy

<400> 3
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Ala Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Ala Gly Ala Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305   310   315   320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325   330   335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340   345   350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355   360   365
Gly Thr Val Ala Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370   375   380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro

385 390 395 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
405 410 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
420 425 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
435 440 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
450 455 460
Ala Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465 470 475 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
485 490 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
500 505 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
515 520 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
530 535 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545 550 555 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
565 570 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Val Cys Gln Ser
580 585 590
Cys Phe

<210> 4
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<400> 4
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95

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

```
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Ala
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 5
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

```
<400> 5
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1               5                   10                  15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50                  55                  60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65                  70                  75                  80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85                  90                  95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100                 105                 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115                 120                 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130                 135                 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145                 150                 155                 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165                 170                 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
```

```
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ala Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Ala Ser Asn Ala Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Ala Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asn Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Ala Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ala Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Val Cys Gln Arg
            580                 585                 590
Cys Leu
```

<210> 6
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<400> 6
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Ala Gly Ala Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300

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

<210> 7
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<400> 7

```
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
```

```
1               5                   10                  15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50                  55                  60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65                  70                  75                  80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85                  90                  95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100                 105                 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115                 120                 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130                 135                 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145                 150                 155                 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165                 170                 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Ala Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Ala Gly Ala Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Ala Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
```

```
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Ala Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 8
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<400> 8

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

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

```
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ala Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Val Cys Gln Arg
            580                 585                 590
Cys Leu
```

<210> 9
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Hyperactive PiggyBac aa sequence

<400> 9

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

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

<210> 10
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac

<220>
<221> SITE
<222> 245..245
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 268..268
<223> Wherein amino acid can be Asp, Asn

<220>
<221> SITE
<222> 275..275
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 277..277
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 287..287
<223> Wherein amino acid can be Ala, Lys

<220>
<221> SITE
<222> 290..290
<223> Wherein amino acid can be Ala, Lys

<220>
<221> SITE
<222> 315..315
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 325..325
<223> Wherein amino acid can be Gly, Ala

<220>

<221> SITE
<222> 341..341
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 346..346
<223> Wherein amino acid can be Asp, Asn

<220>
<221> SITE
<222> 347..347
<223> Wherein amino acid can be Asn, Ala, Ser

<220>
<221> SITE
<222> 350..350
<223> Wherein amino acid can be Thr, Ala

<220>
<221> SITE
<222> 351..351
<223> Xaa can be Ser, Glu, Pro, Ala

<220>
<221> SITE
<222> 356..356
<223> Wherein amino acid can be Lys, Glu

<220>
<221> SITE
<222> 357..357
<223> Wherein amino acid can be Asn, Ala

<220>
<221> SITE
<222> 372..372
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 375..375
<223> Wherein amino acid can be Lys, Ala

<220>
<221> SITE
<222> 388..388
<223> Wherein amino acid can be Arg, Ala

<220>

<221> SITE
<222> 409..409
<223> Wherein amino acid can be Lys, Ala

<220>
<221> SITE
<222> 412..412
<223> Wherein amino acid can be Lys, Ala

<220>
<221> SITE
<222> 432..432
<223> Wherein amino acid can be Lys, Ala

<220>
<221> SITE
<222> 460..460
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 461..461
<223> Wherein amino acid can be Ala, Lys

<220>
<221> SITE
<222> 465..465
<223> Wherein amino acid can be Trp, Ala

<220>
<221> SITE
<222> 560..560
<223> Wherein amino acid can be Thr, Ala

<220>
<221> SITE
<222> 564..564
<223> Wherein amino acid can be Ser, Pro

<220>
<221> SITE
<222> 571..571
<223> Wherein amino acid can be Asn, Ser

<220>
<221> SITE
<222> 573..573
<223> Wherein amino acid can be Ser, Ala

<220>

<221> SITE
<222> 576..576
<223> Wherein amino acid can be Lys, Ala

<220>
<221> SITE
<222> 586..586
<223> Wherein amino acid can be His, any naturally occuring amino acid

<220>
<221> SITE
<222> 587..587
<223> Wherein amino acid can be Ile, Ala

<220>
<221> SITE
<222> 589..589
<223> Wherein amino acid can be Met, Val

<220>
<221> SITE
<222> 592..592
<223> Wherein amino acid can be Ser, Gly

<220>
<221> SITE
<222> 594..594
<223> Wherein amino acid can be Phe, Leu

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

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

Tyr Cys Pro Xaa Lys Ile Arg Arg Lys Ala Xaa Ala Xaa Cys Lys Xaa
565 570 575
Cys Lys Lys Val Ile Cys Arg Glu His Xaa Xaa Asp Xaa Cys Gln Xaa
580 585 590
Cys Xaa

<210> 11
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations in the catalytic triad

<220>
<221> SITE
<222> 268..268
<223> Xaa can be Asp, Asn

<220>
<221> SITE
<222> 346..346
<223> Xaa can be Asp, Asn

<400> 11
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175

```
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Xaa Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Xaa Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Arg Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
```

Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
580 585 590
Cys Phe

<210> 12
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations in amino acids that are critical for excision

<220>
<221> SITE
<222> 287..287
<223> Xaa can be Lys, Ala

<220>
<221> SITE
<222> 290..290
<223> Xaa can be Lys, Ala

<220>
<221> SITE
<222> 460..460
<223> Xaa can be Arg, Ala

<220>
<221> SITE
<222> 461..461
<223> Xaa can be Lys, Ala

<400> 12
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110

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

```
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 13
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutation that are involved in target joining

<220>
<221> SITE
<222> 351..351
<223> Xaa can be Ser, Glu, Pro, Ala

<220>
<221> SITE
<222> 356..356
<223> Xaa can be Lys, Glu

<400> 13

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

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

```
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 14
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations that are critical for integration

<220>
<221> SITE
<222> 560..560
<223> Wherein amino acid can be Thr, Ala,

<220>
<221> SITE
<222> 564..564
<223> Wherein amino acid can be Ser, Pro

<220>
<221> SITE
<222> 571..571
<223> Wherein amino acid can be Asn, Ser

<220>
<221> SITE
<222> 573..573
<223> Wherein amino acid can be Ser, Ala

<220>
<221> SITE
<222> 589..589
<223> Wherein amino acid can be Met, Val

<220>
<221> SITE
<222> 592..592
<223> Wherein amino acid can be Ser, Gly

<220>
<221> SITE
<222> 594..594
<223> Wherein amino acid can be Phe, Leu

<400> 14
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305   310   315   320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325   330   335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile

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


<210> 15
<211> 594
<212> PRT
<213> Artificial Sequence


<220>
<223> Modified hyperactive PiggyBac with mutations that are involved in
      alignment

<220>
<221> SITE
<222> 325..325
<223> Xaa can be Gly, Ala

<220>
```

<221> SITE
<222> 347..347
<223> Xaa can be Asn, Ala, Ser

<220>
<221> SITE
<222> 350..350
<223> Xaa can be Thr, Ala

<220>
<221> SITE
<222> 465..465
<223> Xaa can be Trp, Ala

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

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

<210> 16
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations at well conserved

amino acids

<220>
<221> SITE
<222> 576..576
<223> Xaa can be lys, Ala

<220>
<221> SITE
<222> 587..587
<223> Xaa can be Ile, Ala

<400> 16
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys

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

<210> 17
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations involved in Zn2+ binding

<220>
<221> SITE
<222> 586..586
<223> Wherein amino acid can be His, any naturally occuring amino acid

<400> 17
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305   310   315   320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325   330   335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340   345   350

```
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Arg Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Xaa Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 18
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac with mutations that are involved in integration

<220>
<221> SITE
<222> 315..315
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE

<222> 341..341
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 372..372
<223> Wherein amino acid can be Arg, Ala

<220>
<221> SITE
<222> 375..375
<223> Wherein amino acid can be Lys, Ala

<400> 18
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285

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

<210> 19
<211> 897
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Corynebacterium ulcerans

<400> 19
Met Thr Asn Ala Val Ala Asn His His Val Leu Trp Ala Lys Phe Asp
1   5   10   15
Asn Val Ser Glu Pro Tyr Pro Leu Leu Ala His Leu Leu Asp Thr Ala
            20   25   30
Thr Ala Ala Thr Cys Leu Phe Asn His Trp Leu Arg Lys Gly Leu Arg
        35   40   45
Asp Arg Leu Ser Thr Glu Leu Gly Pro Asp Ala Glu Lys Ile Leu Gly
    50   55   60
Phe Val Ala Gly Ile His Asp Leu Gly Lys Ala Asn Pro Tyr Phe Gln
65   70   75   80
Ala Gln Arg Arg Asn Lys Lys Glu Glu Trp Ile Thr Leu Arg Asp Ala
                85   90   95
Ile Gln Lys Ala Gly Phe Pro Leu Ser Asn Gly Thr Ser Ala Leu Phe
            100   105   110
Glu Glu Thr Lys Glu Lys Arg Arg His Glu Asn Ile Thr Leu Ser Ile
        115   120   125
Leu Gly Trp Glu Ile Thr Lys Phe Leu Gln Val Lys Asp Val Trp Pro
    130   135   140
Gln Leu Ala Ile Ile Gly His His Gly Asn Phe Ser Ala Pro Gly Phe
145   150   155   160
Leu Ser Asp Glu Asp Asp Leu Glu Asp Ile Glu Asp Ile Phe Asp Asp
                165   170   175
Asn Gly Trp Ser Pro Thr His Glu Leu Leu Val Ser Ser Leu Leu Gln
            180   185   190
Ala Val Gly Leu Glu Lys Gln Pro Glu Ile Lys His Ile Ser Pro Ala
        195   200   205
Ser Ala Ile Leu Ile Ser Gly Leu Val Val Leu Ala Asp Arg Ile Ala
    210   215   220
Ser Gln Ser Glu Met Ala Ser Asp Gly Leu Gln Ala Leu Gln Lys Glu
225   230   235   240
Glu Leu Phe Phe His Gln Pro Glu Lys Trp Ile Ala Asn Arg Lys Ala
                245   250   255
Phe Cys Arg Glu Ile Ile Glu Asn Thr Val Gly Thr Tyr His Pro Trp
            260   265   270
Glu Ser Glu Ala Ala Gly Ile Arg Ala Val Leu Gly Asp Tyr Glu Pro
        275   280   285
Arg Phe Thr Gln Lys Ala Ala Leu Asn Ala Gly Asp Gly Leu Phe Asn
    290   295   300
Val Met Glu Thr Thr Gly Ala Gly Lys Thr Glu Ala Ala Leu Leu Arg
305   310   315   320
His Val Lys Arg Lys Glu Arg Leu Leu Phe Phe Leu Pro Thr Gln Ala
                325   330   335
Thr Thr Asn Ala Ile Met Asp Arg Ile Gly Lys Ile Phe Asp Gly Thr
            340   345   350
Pro Asn Val Ala Ser Leu Ala His Gly Leu Ala Val Thr Glu Asp Phe
        355   360   365
Tyr Ala His Pro Ile Leu Pro Val Gln Gly Ser Ser Asp Asp Ala Asn
    370   375   380
Tyr Lys Asp Asn Gly Gly Leu Tyr Pro Thr Glu Phe Val Arg Ser Ala

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

785 790 795 800
Leu Ile Asp Tyr Pro Ser Ile Ser Ala Ile Leu Cys Asp Pro Thr Gly
805 810 815
Thr Ile Pro Gly Ala Trp Thr Asp Ser Val Glu Lys Leu Ile Ala Ile
820 825 830
Ser Ala Lys Asp Ser Glu Ser Leu Arg Arg Ala Leu Arg Ala Ser Ile
835 840 845
Ser Ile Pro His Ser Lys Lys Phe Leu Pro Ile Thr Ser Arg Glu Ile
850 855 860
Pro Leu Ser Glu Ala Lys Thr Leu Leu Ser Gly Tyr Ser Ala Val His
865 870 875 880
Ile Gln Pro Asp Glu Tyr Asp Leu Gln Ser Gly Leu Lys Gly Pro Gln
885 890 895
Lys

<210> 20
<211> 876
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Corynebacterium diphtheria

<400> 20
Met Asn Pro His Glu Glu Leu Trp Ala Lys Gln Lys Gly Leu Ala Lys
1 5 10 15
Pro Tyr Pro Leu Leu Ala His Leu Leu Asp Ser Ala Ala Val Ala Gly
20 25 30
Ala Leu Trp Asp His Trp Leu Arg Gln Asp Leu Arg Gln Met Phe Ile
35 40 45
Glu Glu Leu Gly Ser Asn Ala Arg Glu Ile Ile Gln Phe Val Val Gly
50 55 60
Ser His Asp Ile Gly Lys Ala Thr Pro Leu Phe Gln Tyr Gln Lys Ala
65 70 75 80
Gln Lys Gly Glu Val Trp Asp Ser Ile Arg Tyr Ala Ile Asp Arg Thr
85 90 95
Gly Arg Tyr Gln Lys Pro Leu Pro Ser Ser Tyr Leu Val Lys Lys Thr
100 105 110
Ser Gly Gly Pro Asn Arg His Glu Gln Trp Ser Ser Phe Ala Ser Lys
115 120 125
Asn Glu Tyr Leu Lys Pro Ser Ala Ala Ala Lys Glu Asn Trp Ile Gly
130 135 140
Leu Ala Ile Gly Gly His His Gly Arg Phe Glu Pro Val Gly Tyr Gly
145 150 155 160
Arg His Gln Arg Lys Ala Ala Glu Asp Leu Ala Lys Ser Gly Trp Ser
165 170 175
Ala Ala Gln Gln Asp Leu Leu Arg Ala Leu Glu Lys Ala Ser Gly Ile
180 185 190

```
Thr Arg Ala Ser Leu Pro Ser Glu Leu Ser Pro Glu Leu Thr Leu Val
        195                 200                 205
Leu Ser Gly Leu Thr Ile Leu Ala Asp Arg Ile Ser Ser Thr Glu Ser
    210                 215                 220
Phe Val Ile Thr Gly Ala Arg Met Ile Asp Asp Gly Thr Leu His Leu
225                 230                 235                 240
Ala Thr Pro Ile Asp Trp Leu Lys Thr Arg Lys Leu Asp Ser Glu Lys
                245                 250                 255
His Val Ala Lys Thr Val Gly Ile Tyr His Gly Trp Asn Asn His Glu
            260                 265                 270
Ser Ala Ile His Ser Ile Leu Lys Gly Tyr Asp Pro Arg Pro Leu Gln
        275                 280                 285
Thr Ile Ala Leu Gln Asn Gln Val Gly Leu Leu Asn Leu Met Ala Pro
    290                 295                 300
Thr Gly Asn Gly Lys Thr Glu Ala Ala Ile Leu Arg His Ser Leu Lys
305                 310                 315                 320
Glu Asn Asp Arg Leu Ile Phe Leu Leu Pro Thr Gln Ala Thr Ser Asn
                325                 330                 335
Ala Ile Met Arg Arg Val Gln Gly Ile Tyr Ser Asp Thr Pro Asn Ala
            340                 345                 350
Ala Ala Leu Ala His Ser Leu Ala Ser Val Glu Asp Phe Tyr Gln Thr
        355                 360                 365
Pro Leu Ser Val Phe Asp Asp His Tyr Asp Pro Ser Lys Glu Gln Phe
    370                 375                 380
Glu Ser Ser Met Ser Gly Gly Leu Tyr Pro Ser Ser Phe Val Cys Ser
385                 390                 395                 400
Gly Ala Ala Arg Leu Leu Ala Pro Ile Cys Ile Gly Thr Val Asp Gln
                405                 410                 415
Ala Leu Ala Thr Ala Leu Pro Gly Lys Trp Ile His Leu Arg Ile Leu
            420                 425                 430
Ala Leu Ala Asn Ala His Ile Val Ile Asp Glu Val His Thr Leu Asp
        435                 440                 445
His Tyr Gln Thr Ala Leu Leu Glu Asn Ile Leu Pro Ile Leu Ala Lys
    450                 455                 460
Leu Lys Thr Lys Ile Thr Phe Leu Thr Ala Thr Met Pro Ser Trp Gln
465                 470                 475                 480
Arg Thr Lys Leu Leu Thr Ala Tyr Gly Gly Glu Asp Leu Gln Ile Pro
                485                 490                 495
Pro Thr Val Phe Pro Ala Ala Glu Thr Val Leu Pro Gly Gln Phe Asn
            500                 505                 510
Arg Thr Leu Ile Asp Ser Asp Ser Thr Thr Ile Asp Phe Thr Met Glu
        515                 520                 525
Glu Thr Ser Tyr Asp His Leu Val Glu Ser His Val Lys Trp His Gln
    530                 535                 540
Thr Thr Arg Leu Asn Ala Pro His Ala Arg Ile Gly Leu Ile Cys Asn
545                 550                 555                 560
Thr Val Lys Arg Ala Gln Glu Ile Ala Ala Ala Leu Glu Lys Thr Asn
                565                 570                 575
Asp Arg Ile Val Leu Leu His Ser Arg Met Thr Thr Glu His Arg Arg
            580                 585                 590
```

```
Arg Ser Ala Glu Leu Leu Glu Ser Leu Leu Gly Pro Asn Gly Asn Arg
        595                 600                 605
Lys Thr Ile Thr Val Val Gly Thr Gln Ala Ile Glu Ala Ser Leu Asp
    610                 615                 620
Ile Asp Leu Asp Ile Leu Arg Thr Glu Leu Cys Pro Ala Pro Ser Leu
625                 630                 635                 640
Val Gln Arg Ala Gly Arg Val Trp Arg Arg Asn Asp Pro Tyr Arg Ser
                645                 650                 655
Ser Arg Ile Thr Ala Asp His Lys Pro Ile Ser Val Val Phe Ile Ala
            660                 665                 670
Glu Ala Lys Asp Trp Gln Val Leu Pro Tyr Leu Arg Ala Glu Thr Ser
        675                 680                 685
Arg Thr Gln Arg Trp Leu Glu Lys His Asn Gln Met Phe Leu Pro Gln
    690                 695                 700
Met Ala Gln Glu Phe Ile Asp Ala Ala Thr Val Asp Leu Asp Thr Ala
705                 710                 715                 720
Thr Ser Glu Met Asp Leu Asp Ala Leu Ala Leu Met Gly Ile His Leu
                725                 730                 735
Met Lys Ala Asp Gly Ala Lys Ala Arg Ile Gln Asp Val Leu Asn Ser
            740                 745                 750
Asp Ser Lys Val Ser Asp Phe Ala Leu Leu Thr Ser Lys Asn Glu Ile
        755                 760                 765
Asp Glu Ala Gln Thr Arg Leu Ile Glu Glu Gly Thr His Leu Arg Ile
    770                 775                 780
Ile Leu Gly Asp Glu Asn Glu Ser Ile Pro Gly Gly Trp Lys His Gly
785                 790                 795                 800
Leu Ser Ser Leu Leu Lys Leu Lys Ala Ser Asp Arg Glu Ser Leu Arg
                805                 810                 815
Thr Ala Leu Leu Ala Ser Ile Pro Leu Leu Val Ser Glu Lys Gln Lys
            820                 825                 830
Gln Leu Leu Tyr Gln His Asn Leu Val Pro Leu Ser Ser Ser Lys Thr
        835                 840                 845
Val Leu Ala Gly Phe Tyr Phe Leu Pro Lys Ala Gln Asn Phe Tyr Ser
    850                 855                 860
Lys Asn Leu Gly Phe Ile Trp Pro Glu Glu Lys Asp
865                 870                 875
```

<210> 21
<211> 773
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Spiroplasma syrphidicola

<400> 21

```
Met Asn Tyr Lys Lys Leu Ile Leu Gly Leu Asp Leu Gly Ile Ala Ser
1               5                   10                  15
Cys Gly Trp Ala Val Thr Gly Gln Met Glu Asp Gly Asn Trp Val Leu
```

```
            20                  25                  30
Asp Asp Phe Gly Val Arg Leu Phe Gln Thr Pro Glu Asn Ser Lys Asp
        35                  40                  45
Gly Thr Thr Asn Ala Ala Ala Arg Arg Leu Lys Arg Gly Ala Arg Arg
    50                  55                  60
Leu Ile Lys Arg Arg Lys Asn Arg Ile Lys Asp Leu Lys Asn Leu Phe
65                  70                  75                  80
Glu Lys Ile Asn Phe Ile Asn Lys Ala Ser Leu Asp Lys Tyr Ile Asn
                85                  90                  95
Glu His Ser Ala Thr Asn Leu Val Glu Asp Phe Asn Arg His Glu Leu
            100                 105                 110
Tyr Asn Pro Tyr Phe Leu Arg Ser Ile Gly Ile Thr Glu Lys Leu Thr
        115                 120                 125
Arg Glu Glu Leu Val Trp Ser Leu Ile His Ile Ala Asn Arg Arg Gly
    130                 135                 140
Tyr Lys Asn Lys Phe Ala Phe Asp Ile Glu Gly Asp Gly Lys Lys Arg
145                 150                 155                 160
Glu Thr Lys Leu Asp Glu Ala Ile Ser Asn Ala Leu Ile Ser Ser Asn
                165                 170                 175
Leu Thr Ile Ser Gln Glu Ile Val Arg Asn Lys Lys Phe Arg Asp Ala
            180                 185                 190
Lys Asn Lys Lys Ala Leu Leu Val Arg Asn Lys Gly Gly Lys Glu Gly
        195                 200                 205
Glu Asn Asn Phe Gln Phe Leu Phe Ala Arg Asp Asp Tyr Lys Lys Glu
    210                 215                 220
Val Asp Leu Leu Leu Ala Lys Gln Ala Lys Phe Tyr Pro Glu Leu Thr
225                 230                 235                 240
Glu Glu Ile Arg Ala Lys Ala Ala Asp Ile Ile Phe Arg Gln Arg Asp
                245                 250                 255
Phe Glu Asp Gly Pro Gly Pro Lys Lys Gln Glu Leu Arg Glu Ile Tyr
            260                 265                 270
Lys Lys Glu Asn Lys Gln Phe Ser Lys Asn Phe Thr Gln Leu Glu Gly
        275                 280                 285
Arg Cys Thr Phe Leu Arg Glu Leu Ser Val Gly Tyr Lys Ser Ser Ile
    290                 295                 300
Leu Phe Asp Leu Phe His Ile Ile Ser Glu Val Ser Lys Ile Ser Lys
305                 310                 315                 320
Tyr Ile Glu Glu Asn Asp Gln Leu Ala Gln Asp Ile Ile Ser Ser Phe
                325                 330                 335
Leu Tyr Asn Glu Ala Gly Lys Lys Gly Lys Thr Leu Leu Lys Glu Ile
            340                 345                 350
Leu Lys Lys His His Ile Asn Asp Asp Ile Phe Asp Thr Asn Ala Tyr
        355                 360                 365
Lys Asn Ile Asp Phe Lys Thr Asn Tyr Leu Asn Leu Leu Lys Glu Val
    370                 375                 380
Phe Gly Asn Asp Val Leu Lys Asn Leu Ser Leu Asn Arg Leu Glu Asp
385                 390                 395                 400
Asn Ile Tyr His Gln Leu Gly Phe Ile Ile His Thr Asn Ile Thr Pro
                405                 410                 415
Glu Arg Lys Glu Lys Ala Ile Asn Gln Trp Leu Leu Glu Asn Asn Ile
```

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

<210> 22
<211> 1380
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Prevotella intermedia

<400> 22
Met Lys Arg Ile Leu Gly Leu Asp Leu Gly Thr Thr Ser Ile Gly Trp
1   5   10   15
Ala Leu Val Asn Glu Ala Glu Asn Asn Asn Glu Ala Ser Ser Ile Val
            20   25   30
Arg Leu Gly Val Arg Val Asn Pro Leu Thr Val Asp Glu Lys Ser Asn
        35   40   45
Phe Glu Lys Gly Lys Ala Ile Thr Thr Asn Ala Asp Arg Gln Leu Arg
    50   55   60
His Gly Ala Arg Ile Asn Leu Gln Arg Tyr Lys Leu Arg Arg Gln Asn
65   70   75   80
Leu His Asp Cys Leu Gln Lys Gln Gly Trp Leu Gly Thr Glu Ala Met
                85   90   95
Tyr Glu Glu Gly Lys Ala Ser Thr Phe Glu Thr Tyr Lys Leu Arg Ala
            100   105   110
Lys Ala Ala Glu Glu Glu Ile Ser Leu His Glu Phe Ala Arg Val Leu
        115   120   125
Phe Met Leu Asn Lys Lys Arg Gly Tyr Lys Ser Asn Arg Lys Ala Asn
    130   135   140
Asn Lys Glu Asp Gly Gln Leu Phe Asp Gly Met Thr Ile Ala Lys Lys
145   150   155   160
Leu Tyr Glu Glu His Leu Thr Pro Ala Glu Tyr Ser Leu Gln Leu Leu
                165   170   175
Asn Lys Gly Lys Lys Phe Thr Gln Gly Tyr Tyr Arg Ser Asp Leu Asn
            180   185   190
Ala Glu Leu Glu Arg Ile Trp Asp Glu Gln Lys Lys Tyr Tyr Pro Glu
        195   200   205
Ile Leu Thr Asp Glu Phe Lys Gln Gln Leu Glu Gly Lys Thr Lys Thr
    210   215   220
Asn Thr Ser Lys Ile Phe Leu Ala Lys Tyr Gly Ile Tyr Ser Ala Asp
225   230   235   240
Leu Lys Gly Leu Asp Arg Lys Phe Gln Pro Leu Lys Trp Arg Val Glu
                245   250   255
Ala Leu Gln Gln Gln Val Asp Lys Glu Val Leu Ala Phe Val Ile Ser
            260   265   270
Asp Leu Lys Gly Gln Ile Ala Asn Thr Ser Gly Leu Leu Gly Ala Ile
        275   280   285
Ser Asp Arg Ser Lys Glu Leu Tyr Phe Asn Lys Gln Thr Val Gly Gln
    290   295   300
Tyr Leu Trp Ala Ser Leu Glu Glu Asn Pro His Ile Ser Ile Lys Asn
305   310   315   320
Lys Pro Phe Tyr Arg Gln Asp Tyr Leu Asp Glu Phe Glu Lys Ile Trp
                325   330   335
Glu Thr Gln Ala Ala Phe His Lys Gln Leu Thr Pro Glu Leu Lys Gln
            340   345   350

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

```
Thr Lys Asp Lys Leu Phe Ser Lys Asp Phe Asp Ile Glu His Ile Ile
        755                 760                 765
Pro Lys Ala Arg Leu Phe Asp Asp Ser Phe Ser Asn Lys Thr Leu Glu
    770                 775                 780
Ala Arg Asn Ile Asn Leu Glu Lys Ser Asn Lys Thr Ala Phe Asp Phe
785                 790                 795                 800
Ile Lys Glu Lys Tyr Gly Glu Asp Gly Ala Glu Ala Tyr Lys Lys Lys
                805                 810                 815
Leu Asp Met Leu Leu Glu Asn Asp Ala Ile Ser Arg Pro Lys Tyr Asn
            820                 825                 830
Asn Leu Leu Arg Ala Glu Ala Asp Ile Pro Ser Asp Phe Ile Asn Arg
        835                 840                 845
Asp Leu Arg Asn Thr Gln Tyr Ile Ala Lys Lys Ala Cys Glu Ile Leu
    850                 855                 860
Gly Glu Leu Val Lys Thr Val Thr Pro Thr Thr Gly Lys Ile Thr Asn
865                 870                 875                 880
Arg Leu Arg Glu Asp Trp Gln Leu Val Asp Val Met Lys Glu Leu Asn
                885                 890                 895
Phe Glu Lys Tyr Glu Lys Leu Gly Leu Thr Glu Ile Val Glu Asp Arg
            900                 905                 910
Asp Gly Arg Lys Ile Lys Arg Ile Lys Asp Trp Thr Lys Arg Asn Asp
        915                 920                 925
His Arg His His Ala Met Asp Ala Leu Ala Ile Ala Phe Thr Lys Pro
    930                 935                 940
Ser Phe Ile Gln Tyr Leu Asn Asn Leu Asn Ala Arg Ser Asn Lys Gly
945                 950                 955                 960
Asp Ser Ile Tyr Ala Ile Glu Asn Lys Glu Leu His Tyr Glu Glu Gly
                965                 970                 975
Lys Leu Arg Phe Asn Ala Pro Ile Pro Val Asn Glu Phe Arg Ala Glu
            980                 985                 990
Ala Lys Arg His Leu Ser Ala Ile Leu Val Ser Ile Lys Ala Lys Asn
        995                 1000                1005
Lys Val Met Thr Gln Asn Val Asn Lys Ile Lys Thr Lys His Gly Ile
    1010                1015                1020
Ile Lys Lys Ile Gln Leu Thr Pro Arg Gly Pro Leu His Asn Glu Thr
1025                1030                1035                1040
Ile Tyr Gly Thr Lys Met Arg Pro Ile Ile Lys Met Val Lys Val Gly
                1045                1050                1055
Ala Ala Leu Asp Glu Ala Thr Ile Asn Lys Val Ser Ser Pro Ala Ile
            1060                1065                1070
Arg Glu Ala Leu Leu Lys Arg Leu Asn Glu Tyr Ser Gly Asn Ala Lys
        1075                1080                1085
Lys Ala Phe Thr Gly Lys Asn Thr Leu Glu Lys Asn Pro Ile Tyr Leu
    1090                1095                1100
Asn Ala Gly Arg Thr Lys Thr Val Pro Ser Leu Val Lys Thr Val Glu
1105                1110                1115                1120
Trp Glu Ser Phe His Pro Thr Arg Lys Leu Ile Asp Lys Asp Leu Asn
                1125                1130                1135
Val Asp Lys Val Val Asp Lys Gly Ile Arg Glu Ile Leu Lys Ala Arg
            1140                1145                1150
```

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

<210> 23
<211> 535
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Spiroplasma taiwanense

<400> 23

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

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

```
465                 470                 475                 480
Val Ile Glu Met Ala Lys Lys Asn Ser Ala Glu Lys Lys Gln Phe Ile
                485                 490                 495
Ser Lys Ile Gln Arg Gln Asn Val Asp Leu Val Lys Lys Leu Ser Asn
            500                 505                 510
Asp Tyr Ser Leu Asp Glu Asn Lys Leu Asn Phe Lys Met Lys Glu Lys
        515                 520                 525
Phe Leu Leu Leu Ser Glu Gln
    530                 535
```

<210> 24
<211> 1281
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Streptococcus iniae

<400> 24

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

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

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

Asn Leu Met Asn Phe Phe Lys Lys Glu Ile Lys Leu Ala Asp Asp Thr
1045 1050 1055
Ile Phe Thr Arg Pro Gln Ile Glu Val Asn Thr Glu Thr Gly Glu Ile
1060 1065 1070
Val Trp Asp Lys Val Lys Asp Met Gln Thr Ile Arg Lys Val Met Ser
1075 1080 1085
Tyr Pro Gln Val Asn Ile Val Met Lys Thr Glu Val Gln Thr Gly Gly
1090 1095 1100
Phe Ser Lys Glu Ser Ile Trp Pro Lys Gly Asp Ser Asp Lys Leu Ile
1105 1110 1115 1120
Ala Arg Lys Lys Ser Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
1125 1130 1135
Pro Ile Ile Ala Tyr Ser Val Leu Val Val Ala Lys Ile Ala Lys Gly
1140 1145 1150
Lys Thr Gln Lys Leu Lys Thr Ile Lys Glu Leu Val Gly Ile Lys Ile
1155 1160 1165
Met Glu Gln Asp Glu Phe Glu Lys Asp Pro Ile Ala Phe Leu Glu Lys
1170 1175 1180
Lys Gly Tyr Gln Asp Ile Gln Thr Ser Ser Ile Ile Lys Leu Pro Lys
1185 1190 1195 1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Leu Leu Ala Ser
1205 1210 1215
Ala Lys Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Asn Lys Tyr
1220 1225 1230
Val Lys Phe Leu Tyr Leu Ala Ser His Tyr Thr Lys Phe Thr Gly Lys
1235 1240 1245
Glu Glu Asp Arg Glu Lys Lys Arg Ser Tyr Val Glu Ser His Leu Tyr
1250 1255 1260
Tyr Phe Asp Val Arg Leu Ser Gln Val Phe Arg Val Thr Asn Val Glu
1265 1270 1275 1280
Phe

<210> 25
<211> 1352
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Belliella baltica

<400> 25
Met Lys Lys Ile Leu Gly Leu Asp Leu Gly Thr Thr Ser Ile Gly Trp
1 5 10 15
Ala Phe Ile Lys Glu Pro Glu Lys Asp Val Val Gly Ser Glu Ile Val
20 25 30
Asp Met Gly Val Arg Ile Val Pro Leu Ser Ser Asp Glu Glu Asn Asp
35 40 45
Phe Ala Lys Gly Asn Thr Ile Ser Ile Asn Ala Asp Arg Thr Leu Lys

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

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

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

```
    1250                1255                1260
Phe Val Pro Ala Asp Val Asp Leu Met Asp Glu Gln Asn Tyr His Leu
1265                1270                1275                1280
Ile Ser Pro Asn Leu Phe Arg Val Gln Lys Ile Ala Arg Lys Asn Tyr
                1285                1290                1295
Val Phe Asn Asn His Leu Glu Thr Lys Ala Val Asp Asn Asp Leu Leu
            1300                1305                1310
Lys Ser Lys Lys Glu Leu Ser Lys Ile Thr Tyr His Phe Tyr Gln Thr
        1315                1320                1325
Pro Glu His Leu Arg Gly Ile Ile Lys Ile Arg Ile Asn His Leu Gly
    1330                1335                1340
Lys Ile Ile Gln Ile Gly Glu Tyr
1345                1350
```

<210> 26
<211> 1509
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Psychroflexus torquisI

<400> 26

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

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

```
Pro Phe Val Ser Asp Tyr Ala Ser Tyr Ser Lys Lys Ala Ile Ser Lys
    610                 615                 620
Leu Leu Pro Ile Met Arg Met Gly Lys Tyr Trp Ser Glu Ser Ala Val
625                 630                 635                 640
Pro Thr Gln Val Lys Glu Arg Ser Leu Ser Ile Met Glu Arg Val Lys
                645                 650                 655
Val Leu Pro Leu Lys Glu Gly Tyr Ser Asp Lys Asp Leu Ala Asp Leu
            660                 665                 670
Leu Ser Arg Val Ser Asp Asp Asp Ile Pro Lys Gln Leu Ile Lys Ser
        675                 680                 685
Phe Ile Ser Phe Lys Asp Lys Asn Pro Leu Lys Gly Leu Asn Thr Tyr
    690                 695                 700
Gln Ala Asn Tyr Leu Val Tyr Gly Arg His Ser Glu Thr Gly Asp Ile
705                 710                 715                 720
Gln His Trp Lys Thr Pro Glu Asp Ile Asp Arg Tyr Leu Asn Asn Phe
                725                 730                 735
Lys Gln His Ser Leu Arg Asn Pro Ile Val Glu Gln Val Val Met Glu
            740                 745                 750
Thr Leu Arg Val Val Arg Asp Ile Trp Glu His Tyr Gly Asn Asn Glu
        755                 760                 765
Lys Asp Phe Phe Lys Glu Ile His Val Glu Leu Gly Arg Glu Met Lys
    770                 775                 780
Ser Pro Ala Gly Lys Arg Glu Lys Leu Ser Gln Arg Asn Thr Glu Asn
785                 790                 795                 800
Glu Asn Thr Asn His Arg Ile Arg Glu Val Leu Lys Glu Leu Met Asn
                805                 810                 815
Asp Ala Ser Val Glu Gly Gly Val Arg Asp Tyr Ser Pro Ser Gln Gln
            820                 825                 830
Glu Ile Leu Lys Leu Tyr Glu Glu Gly Ile Tyr Gln Asn Pro Asn Thr
        835                 840                 845
Asn Tyr Leu Lys Val Asp Glu Asp Glu Ile Leu Lys Ile Arg Lys Lys
    850                 855                 860
Asn Asn Pro Thr Gln Lys Glu Ile Gln Arg Tyr Lys Leu Trp Leu Glu
865                 870                 875                 880
Gln Gly Tyr Ile Ser Pro Tyr Thr Gly Lys Ile Ile Pro Leu Thr Lys
                885                 890                 895
Leu Phe Thr His Glu Tyr Gln Ile Glu His Ile Ile Pro Gln Ser Arg
            900                 905                 910
Tyr Tyr Asp Asn Ser Leu Gly Asn Lys Ile Ile Cys Glu Ser Glu Val
        915                 920                 925
Asn Glu Asp Lys Asp Asn Lys Thr Ala Tyr Glu Tyr Leu Lys Val Glu
    930                 935                 940
Lys Gly Ser Ile Val Phe Gly His Lys Leu Leu Asn Leu Asp Glu Tyr
945                 950                 955                 960
Glu Ala His Val Asn Lys Tyr Phe Lys Lys Asn Lys Thr Lys Leu Lys
                965                 970                 975
Asn Leu Leu Ser Glu Asp Ile Pro Glu Gly Phe Ile Asn Arg Gln Leu
            980                 985                 990
Asn Asp Ser Arg Tyr Ile Ser Lys Leu Val Lys Gly Leu Leu Ser Asn
        995                 1000                1005
```

```
Ile Val Arg Glu Asn Gly Glu Gln Glu Ala Thr Ser Lys Asn Leu Ile
    1010                1015                1020
Pro Val Thr Gly Val Val Thr Ser Lys Leu Lys Gln Asp Trp Gly Leu
1025                1030                1035                1040
Asn Asp Lys Trp Asn Glu Ile Ile Ala Pro Arg Phe Lys Arg Leu Asn
                1045                1050                1055
Lys Leu Thr Asn Ser Asn Asp Phe Gly Phe Trp Asp Asn Asp Ile Asn
            1060                1065                1070
Ala Phe Arg Ile Gln Val Pro Asp Ser Leu Ile Lys Gly Phe Ser Lys
        1075                1080                1085
Lys Arg Ile Asp His Arg His His Ala Leu Asp Ala Leu Val Val Ala
    1090                1095                1100
Cys Thr Ser Arg Asn His Thr His Tyr Leu Ser Ala Leu Asn Ala Glu
1105                1110                1115                1120
Asn Lys Asn Tyr Ser Leu Arg Asp Lys Leu Val Ile Lys Asn Glu Asn
                1125                1130                1135
Gly Asp Tyr Thr Lys Thr Phe Gln Ile Pro Trp Gln Gly Phe Thr Ile
            1140                1145                1150
Glu Ala Lys Asn Asn Leu Glu Lys Thr Val Val Ser Phe Lys Lys Asn
        1155                1160                1165
Leu Arg Val Ile Asn Lys Thr Asn Asn Lys Phe Trp Ser Tyr Lys Asp
    1170                1175                1180
Glu Asn Gly Asn Leu Asn Leu Gly Lys Asp Gly Lys Pro Lys Lys Lys
1185                1190                1195                1200
Leu Arg Lys Gln Thr Lys Gly Tyr Asn Trp Ala Ile Arg Lys Pro Leu
                1205                1210                1215
His Lys Glu Thr Val Ser Gly Ile Tyr Asn Ile Asn Ala Pro Lys Asn
            1220                1225                1230
Lys Ile Ala Thr Ser Val Arg Thr Leu Leu Thr Glu Ile Lys Asn Glu
        1235                1240                1245
Lys His Leu Ala Lys Ile Thr Asp Leu Arg Ile Arg Glu Thr Ile Leu
    1250                1255                1260
Pro Asn His Leu Lys His Tyr Leu Asn Asn Lys Gly Glu Ala Asn Phe
1265                1270                1275                1280
Ser Glu Ala Phe Ser Gln Gly Gly Ile Glu Asp Leu Asn Lys Lys Ile
                1285                1290                1295
Thr Thr Leu Asn Glu Gly Lys Lys His Gln Pro Ile Tyr Arg Val Lys
            1300                1305                1310
Ile Phe Glu Val Gly Ser Lys Phe Ser Ile Ser Glu Asp Glu Asn Ser
        1315                1320                1325
Ala Lys Ser Lys Lys Tyr Val Glu Ala Ala Lys Gly Thr Asn Leu Phe
    1330                1335                1340
Phe Ala Ile Tyr Leu Asp Glu Glu Asn Lys Lys Arg Asn Tyr Glu Thr
1345                1350                1355                1360
Ile Pro Leu Asn Glu Val Ile Thr His Gln Lys Gln Val Ala Gly Phe
                1365                1370                1375
Pro Lys Ser Glu Arg Leu Ser Val Gln Pro Asp Ser Gln Lys Gly Thr
            1380                1385                1390
Phe Leu Phe Thr Leu Ser Pro Asn Asp Leu Val Tyr Val Pro Asn Asn
        1395                1400                1405
```

Glu Glu Leu Glu Asn Arg Asp Leu Phe Asn Leu Gly Asn Leu Asn Val
    1410   1415   1420
Glu Gln Ile Ser Arg Ile Tyr Lys Phe Thr Asp Ser Ser Asp Lys Thr
1425   1430   1435   1440
Cys Asn Phe Ile Pro Phe Gln Val Ser Lys Leu Ile Phe Asn Leu Lys
                1445   1450   1455
Lys Lys Glu Gln Lys Lys Leu Asp Val Asp Phe Ile Ile Gln Asn Glu
            1460   1465   1470
Phe Gly Leu Gly Ser Pro Gln Ser Lys Asn Gln Lys Ser Ile Asp Asp
        1475   1480   1485
Val Met Ile Lys Glu Lys Cys Ile Lys Leu Lys Ile Asp Arg Leu Gly
    1490   1495   1500
Asn Ile Ser Lys Ala
1505

<210> 27
<211> 1388
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Streptococcus thermophilus

<400> 27
Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1   5   10   15
Gly Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met
            20   25   30
Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu
        35   40   45
Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu
    50   55   60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
65   70   75   80
Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala
                85   90   95
Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg
            100   105   110
Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr
        115   120   125
His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp
    130   135   140
Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145   150   155   160
Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser
                165   170   175
Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr
            180   185   190
Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu

```
        195                 200                 205
Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg
    210                 215                 220
Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu
225                 230                 235                 240
Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe
                245                 250                 255
Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp
            260                 265                 270
Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp
        275                 280                 285
Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
    290                 295                 300
Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala
305                 310                 315                 320
Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys
                325                 330                 335
Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys
            340                 345                 350
Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
        355                 360                 365
Gln Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu
    370                 375                 380
Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg
385                 390                 395                 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu
                405                 410                 415
Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe
            420                 425                 430
Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp
    450                 455                 460
Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp
465                 470                 475                 480
Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr
                485                 490                 495
Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu Leu Thr Lys Val Arg
        515                 520                 525
Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe Leu Asp Ser Lys Gln
    530                 535                 540
Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp Lys Arg Lys Val Thr
545                 550                 555                 560
Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly
                565                 570                 575
Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr
            580                 585                 590
Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp
```

```
        595                 600                 605
Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile
    610                 615                 620
Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn
625                 630                 635                 640
Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr
                645                 650                 655
Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu
            660                 665                 670
Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser
        675                 680                 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys
    690                 695                 700
Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn
705                 710                 715                 720
Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys
                725                 730                 735
Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met
            740                 745                 750
Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn
        755                 760                 765
Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg
    770                 775                 780
Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn
785                 790                 795                 800
Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp
                805                 810                 815
Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly
            820                 825                 830
Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile
        835                 840                 845
Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu
    850                 855                 860
Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu
865                 870                 875                 880
Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser
                885                 890                 895
Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
            900                 905                 910
Gly Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu
        915                 920                 925
Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu
    930                 935                 940
Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val
945                 950                 955                 960
Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp
                965                 970                 975
Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His
            980                 985                 990
Asp Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr
```

```
        995                 1000                1005
Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn
    1010                1015                1020
Ser Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser
1025                1030                1035                1040
Asn Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg
                1045                1050                1055
Val Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser
            1060                1065                1070
Val Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser
        1075                1080                1085
Tyr Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly
    1090                1095                1100
Leu Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser
1105                1110                1115                1120
Lys Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr
                1125                1130                1135
Leu Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe
            1140                1145                1150
Thr Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile
        1155                1160                1165
Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn
    1170                1175                1180
Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp
1185                1190                1195                1200
Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser
                1205                1210                1215
Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys
            1220                1225                1230
Arg Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe
        1235                1240                1245
Val Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu
    1250                1255                1260
Asn His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu Glu Leu
1265                1270                1275                1280
Phe Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly Ala Lys Lys
                1285                1290                1295
Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser
            1300                1305                1310
Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg
        1315                1320                1325
Lys Gly Leu Phe Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu
    1330                1335                1340
Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser
1345                1350                1355                1360
Leu Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr
                1365                1370                1375
Glu Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
            1380                1385
```

<210> 28
<211> 1334
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Listeria innocua

<400> 28
Met Lys Lys Pro Tyr Thr Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Leu Thr Asp Gln Tyr Asp Leu Val Lys Arg Lys Met
            20 25 30
Lys Ile Ala Gly Asp Ser Glu Lys Lys Gln Ile Lys Lys Asn Phe Trp
        35 40 45
Gly Val Arg Leu Phe Asp Glu Gly Gln Thr Ala Ala Asp Arg Arg Met
    50 55 60
Ala Arg Thr Ala Arg Arg Arg Ile Glu Arg Arg Arg Asn Arg Ile Ser
65 70 75 80
Tyr Leu Gln Gly Ile Phe Ala Glu Glu Met Ser Lys Thr Asp Ala Asn
                85 90 95
Phe Phe Cys Arg Leu Ser Asp Ser Phe Tyr Val Asp Asn Glu Lys Arg
            100 105 110
Asn Ser Arg His Pro Phe Phe Ala Thr Ile Glu Glu Glu Val Glu Tyr
        115 120 125
His Lys Asn Tyr Pro Thr Ile Tyr His Leu Arg Glu Glu Leu Val Asn
    130 135 140
Ser Ser Glu Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145 150 155 160
Ile Ile Lys Tyr Arg Gly Asn Phe Leu Ile Glu Gly Ala Leu Asp Thr
                165 170 175
Gln Asn Thr Ser Val Asp Gly Ile Tyr Lys Gln Phe Ile Gln Thr Tyr
            180 185 190
Asn Gln Val Phe Ala Ser Gly Ile Glu Asp Gly Ser Leu Lys Lys Leu
        195 200 205
Glu Asp Asn Lys Asp Val Ala Lys Ile Leu Val Glu Lys Val Thr Arg
    210 215 220
Lys Glu Lys Leu Glu Arg Ile Leu Lys Leu Tyr Pro Gly Glu Lys Ser
225 230 235 240
Ala Gly Met Phe Ala Gln Phe Ile Ser Leu Ile Val Gly Ser Lys Gly
                245 250 255
Asn Phe Gln Lys Pro Phe Asp Leu Ile Glu Lys Ser Asp Ile Glu Cys
            260 265 270
Ala Lys Asp Ser Tyr Glu Glu Asp Leu Glu Ser Leu Leu Ala Leu Ile
        275 280 285
Gly Asp Glu Tyr Ala Glu Leu Phe Val Ala Ala Lys Asn Ala Tyr Ser
    290 295 300
Ala Val Val Leu Ser Ser Ile Ile Thr Val Ala Glu Thr Glu Thr Asn
305 310 315 320

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

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

Ala Tyr Ala Val Val Ile Glu Tyr Ala Lys Gly Lys Asn Lys Leu Val
1125 1130 1135
Phe Glu Lys Lys Ile Ile Arg Val Thr Ile Met Glu Arg Lys Ala Phe
1140 1145 1150
Glu Lys Asp Glu Lys Ala Phe Leu Glu Glu Gln Gly Tyr Arg Gln Pro
1155 1160 1165
Lys Val Leu Ala Lys Leu Pro Lys Tyr Thr Leu Tyr Glu Cys Glu Glu
1170 1175 1180
Gly Arg Arg Arg Met Leu Ala Ser Ala Asn Glu Ala Gln Lys Gly Asn
1185 1190 1195 1200
Gln Gln Val Leu Pro Asn His Leu Val Thr Leu Leu His His Ala Ala
1205 1210 1215
Asn Cys Glu Val Ser Asp Gly Lys Ser Leu Asp Tyr Ile Glu Ser Asn
1220 1225 1230
Arg Glu Met Phe Ala Glu Leu Leu Ala His Val Ser Glu Phe Ala Lys
1235 1240 1245
Arg Tyr Thr Leu Ala Glu Ala Asn Leu Asn Lys Ile Asn Gln Leu Phe
1250 1255 1260
Glu Gln Asn Lys Glu Gly Asp Ile Lys Ala Ile Ala Gln Ser Phe Val
1265 1270 1275 1280
Asp Leu Met Ala Phe Asn Ala Met Gly Ala Pro Ala Ser Phe Lys Phe
1285 1290 1295
Phe Glu Thr Thr Ile Glu Arg Lys Arg Tyr Asn Asn Leu Lys Glu Leu
1300 1305 1310
Leu Asn Ser Thr Ile Ile Tyr Gln Ser Ile Thr Gly Leu Tyr Glu Ser
1315 1320 1325
Arg Lys Arg Leu Asp Asp
1330

<210> 29
<211> 984
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Campylobacter jejuni

<400> 29
Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
1 5 10 15
Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
20 25 30
Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
35 40 45
Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg
50 55 60
Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
65 70 75 80
Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly

```
                85                  90                  95
Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu
            100                 105                 110
Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg
        115                 120                 125
Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala
    130                 135                 140
Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln
145                 150                 155                 160
Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu
                165                 170                 175
Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu
            180                 185                 190
Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe
        195                 200                 205
Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu
    210                 215                 220
Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser
225                 230                 235                 240
His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro
                245                 250                 255
Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile
            260                 265                 270
Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys
        275                 280                 285
Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu
    290                 295                 300
Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu
305                 310                 315                 320
Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys
                325                 330                 335
Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu
            340                 345                 350
Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
        355                 360                 365
Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
    370                 375                 380
Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
385                 390                 395                 400
Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
                405                 410                 415
Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
            420                 425                 430
Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr
        435                 440                 445
Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn
    450                 455                 460
Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
465                 470                 475                 480
Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
```

```
                485                 490                 495
Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys
            500                 505                 510
Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
        515                 520                 525
Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
    530                 535                 540
Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
545                 550                 555                 560
Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu
                565                 570                 575
Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
            580                 585                 590
Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala
        595                 600                 605
Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
    610                 615                 620
Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
625                 630                 635                 640
Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp
                645                 650                 655
Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln
            660                 665                 670
Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser
        675                 680                 685
Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His
    690                 695                 700
Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser
705                 710                 715                 720
Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser
                725                 730                 735
Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys
            740                 745                 750
Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp
        755                 760                 765
Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser
    770                 775                 780
Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln
785                 790                 795                 800
Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys
                805                 810                 815
Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg
            820                 825                 830
Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro
        835                 840                 845
Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
    850                 855                 860
Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865                 870                 875                 880
Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
```

```
                885                 890                 895
Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
            900                 905                 910
Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
        915                 920                 925
Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
    930                 935                 940
Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
945                 950                 955                 960
Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
                965                 970                 975
Arg Gln Arg Glu Asp Phe Lys Lys
            980
```

<210> 30
<211> 1082
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Neisseria. meningitidis

<400> 30

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

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

```
Ser Glu Asn Gln Asn Lys Gly Asn Gln Thr Pro Tyr Glu Tyr Phe Asn
    610                 615                 620
Gly Lys Asp Asn Ser Arg Glu Trp Gln Glu Phe Lys Ala Arg Val Glu
625                 630                 635                 640
Thr Ser Arg Phe Pro Arg Ser Lys Lys Gln Arg Ile Leu Leu Gln Lys
                645                 650                 655
Phe Asp Glu Asp Gly Phe Lys Glu Arg Asn Leu Asn Asp Thr Arg Tyr
            660                 665                 670
Val Asn Arg Phe Leu Cys Gln Phe Val Ala Asp Arg Met Arg Leu Thr
        675                 680                 685
Gly Lys Gly Lys Lys Arg Val Phe Ala Ser Asn Gly Gln Ile Thr Asn
    690                 695                 700
Leu Leu Arg Gly Phe Trp Gly Leu Arg Lys Val Arg Ala Glu Asn Asp
705                 710                 715                 720
Arg His His Ala Leu Asp Ala Val Val Val Ala Cys Ser Thr Val Ala
                725                 730                 735
Met Gln Gln Lys Ile Thr Arg Phe Val Arg Tyr Lys Glu Met Asn Ala
            740                 745                 750
Phe Asp Gly Lys Thr Ile Asp Lys Glu Thr Gly Glu Val Leu His Gln
        755                 760                 765
Lys Thr His Phe Pro Gln Pro Trp Glu Phe Phe Ala Gln Glu Val Met
    770                 775                 780
Ile Arg Val Phe Gly Lys Pro Asp Gly Lys Pro Glu Phe Glu Glu Ala
785                 790                 795                 800
Asp Thr Pro Glu Lys Leu Arg Thr Leu Leu Ala Glu Lys Leu Ser Ser
                805                 810                 815
Arg Pro Glu Ala Val His Glu Tyr Val Thr Pro Leu Phe Val Ser Arg
            820                 825                 830
Ala Pro Asn Arg Lys Met Ser Gly Gln Gly His Met Glu Thr Val Lys
        835                 840                 845
Ser Ala Lys Arg Leu Asp Glu Gly Val Ser Val Leu Arg Val Pro Leu
    850                 855                 860
Thr Gln Leu Lys Leu Lys Asp Leu Glu Lys Met Val Asn Arg Glu Arg
865                 870                 875                 880
Glu Pro Lys Leu Tyr Glu Ala Leu Lys Ala Arg Leu Glu Ala His Lys
                885                 890                 895
Asp Asp Pro Ala Lys Ala Phe Ala Glu Pro Phe Tyr Lys Tyr Asp Lys
            900                 905                 910
Ala Gly Asn Arg Thr Gln Gln Val Lys Ala Val Arg Val Glu Gln Val
        915                 920                 925
Gln Lys Thr Gly Val Trp Val Arg Asn His Asn Gly Ile Ala Asp Asn
    930                 935                 940
Ala Thr Met Val Arg Val Asp Val Phe Glu Lys Gly Asp Lys Tyr Tyr
945                 950                 955                 960
Leu Val Pro Ile Tyr Ser Trp Gln Val Ala Lys Gly Ile Leu Pro Asp
                965                 970                 975
Arg Ala Val Val Gln Gly Lys Asp Glu Glu Asp Trp Gln Leu Ile Asp
            980                 985                 990
Asp Ser Phe Asn Phe Lys Phe Ser Leu His Pro Asn Asp Leu Val Glu
        995                 1000                1005
```

Val Ile Thr Lys Lys Ala Arg Met Phe Gly Tyr Phe Ala Ser Cys His
1010 1015 1020
Arg Gly Thr Gly Asn Ile Asn Ile Arg Ile His Asp Leu Asp His Lys
1025 1030 1035 1040
Ile Gly Lys Asn Gly Ile Leu Glu Gly Ile Gly Val Lys Thr Ala Leu
1045 1050 1055
Ser Phe Gln Lys Tyr Gln Ile Asp Glu Leu Gly Lys Glu Ile Arg Pro
1060 1065 1070
Cys Arg Leu Lys Lys Arg Pro Pro Val Arg
1075 1080

<210> 31
<211> 1367
<212> PRT
<213> Artificial Sequence

<220>
<223> Cas9 from Streptococcus pyogenes

<400> 31
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn

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

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

1025 1030 1035 1040
Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile
1045 1050 1055
Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val
1060 1065 1070
Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met
1075 1080 1085
Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe
1090 1095 1100
Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala
1105 1110 1115 1120
Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1125 1130 1135
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys
1140 1145 1150
Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met
1155 1160 1165
Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys
1170 1175 1180
Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr
1185 1190 1195 1200
Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala
1205 1210 1215
Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His Tyr
1250 1255 1260
Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val Ile
1265 1270 1275 1280
Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys His
1285 1290 1295
Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe
1300 1305 1310
Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr
1315 1320 1325
Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala
1330 1335 1340
Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp
1345 1350 1355 1360
Leu Ser Gln Leu Gly Gly Asp
1365

<210> 32
<211> 528
<212> DNA
<213> Artificial Sequence

<220>
<223> zfp coding sequence

<400> 32
atggcccagg ctgctcttga gcccggagag aaaccctaca agtgcccgga gtgcggaaag 60

tccttctctg agcggagtca cctccgagag caccagcgga ctcatacggg cgaaaaacca 120

tacaagtgcc cagaatgtgg taaatctttt tctcgggctg acaacctgac tgaacatcag 180

cgcacgcaca ccggtgaaaa accttacaag tgtccagagt gtggcaagag cttttctagt 240

agaaggacct gtcgagcgca tcagcggact cacaccggcg aaaaacccta taagtgtccg 300

gaatgtggaa agagctttag ccgcaacgac acccttactg aacaccagcg aacacacacg 360

ggagaaaaac catataaatg tccggaatgt ggcaaaagtt ttagtcggag tgataaactt 420

acggagcacc aacggacaca caccggagag aagccatata agtgtcctga atgtggaaag 480

tccttctcac agcttgctca tctgcgagca catcagcgca cacacacc 528

<210> 33
<211> 176
<212> PRT
<213> Artificial Sequence

<220>
<223> ZFP amino acid sequence

<400> 33
Met Ala Gln Ala Ala Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro
1 5 10 15
Glu Cys Gly Lys Ser Phe Ser Glu Arg Ser His Leu Arg Glu His Gln
20 25 30
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
35 40 45
Ser Phe Ser Arg Ala Asp Asn Leu Thr Glu His Gln Arg Thr His Thr
50 55 60
Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Ser
65 70 75 80
Arg Arg Thr Cys Arg Ala His Gln Arg Thr His Thr Gly Glu Lys Pro
85 90 95
Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Asn Asp Thr Leu
100 105 110
Thr Glu His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro
115 120 125
Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Lys Leu Thr Glu His Gln

```
    130                 135                 140
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
145                 150                 155                 160
Ser Phe Ser Gln Leu Ala His Leu Arg Ala His Gln Arg Thr His Thr
                165                 170                 175
```

<210> 34
<211> 555
<212> DNA
<213> Artificial Sequence

<220>
<223> ZNF-E2C Nucleotide

<400> 34

```
atggcgcagg cggcgctgga accgggcgaa aaaccgtata aatgcccgga atgcggcaaa      60

agctttagcc gcaaagatag cctggtgcgc catcagcgca cccataccgg cgaaaaaccg     120

tataaatgcc cggaatgcgg caaaagcttt agccagagcg gcgatctgcg ccgccatcag     180

cgcacccata ccggcgaaaa accgtataaa tgcccggaat gcggcaaaag ctttagcgat     240

tgccgcgatc tggcgcgcca tcagcgcacc ataccggcg aaaaaccgta taaatgcccg     300

gaatgcggca aaagctttag ccagagcagc catctggtgc gccatcagcg cacccatacc     360

ggcgaaaaac cgtataaatg cccggaatgc ggcaaaagct ttagcgattg ccgcgatctg     420

gcgcgccatc agcgcaccca taccggcgaa aaaccgtata aatgcccgga atgcggcaaa     480

agctttagcc gcagcgataa actggtgcgc catcagcgca cccataccgg caaaaaaacc     540

agcggccagg cgggc                                                      555
```

<210> 35
<211> 185
<212> PRT
<213> Artificial Sequence

<220>
<223> ZNF-E2C Aminoacid

<400> 35

```
Met Ala Gln Ala Ala Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro
```

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

<210> 36
<211> 555
<212> DNA
<213> Artificial Sequence

<220>
<223> ZNF-E3 Nucleotide

<400> 36
atggcgcagg cggcgctgga accgggcgaa aaaccgtata aatgcccgga atgcggcaaa 60

agctttagcg atccgggcgc gctggtgcgc catcagcgca cccataccgg cgaaaaaccg 120

tataaatgcc cggaatgcgg caaaagcttt agccagagca gccatctggt gcgccatcag 180

cgcacccata ccggcgaaaa accgtataaa tgcccggaat gcggcaaaag ctttagcgat 240

tgccgcgatc tggcgcgcca tcagcgcacc ataccggcg aaaaaccgta taaatgcccg 300

gaatgcggca aaagctttag ccagagcagc catctggtgc gccatcagcg cacccatacc 360

ggcgaaaaac cgtataaatg cccggaatgc ggcaaaagct ttagcgattg ccgcgatctg 420

gcgcgccatc agcgcaccca taccggcgaa aaaccgtata aatgcccgga atgcggcaaa 480

agctttagcc agagcagcca tctggtgcgc catcagcgca cccataccgg caaaaaaacc 540

agcggccagg cgggc 555

<210> 37
<211> 185
<212> PRT
<213> Artificial Sequence

<220>
<223> ZNF-E3 Amino acid

<400> 37
Met Ala Gln Ala Ala Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro
1 5 10 15
Glu Cys Gly Lys Ser Phe Ser Asp Pro Gly Ala Leu Val Arg His Gln
20 25 30
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
35 40 45
Ser Phe Ser Gln Ser Ser His Leu Val Arg His Gln Arg Thr His Thr
50 55 60
Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp
65 70 75 80
Cys Arg Asp Leu Ala Arg His Gln Arg Thr His Thr Gly Glu Lys Pro
85 90 95
Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser His Leu
100 105 110
Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro
115 120 125
Glu Cys Gly Lys Ser Phe Ser Asp Cys Arg Asp Leu Ala Arg His Gln
130 135 140
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
145 150 155 160
Ser Phe Ser Gln Ser Ser His Leu Val Arg His Gln Arg Thr His Thr
165 170 175
Gly Lys Lys Thr Ser Gly Gln Ala Gly
180 185

<210> 38
<211> 528
<212> DNA
<213> Artificial Sequence

<220>
<223> ZNF-TRCa Nucleotide

<400> 38

atggcgcagg cggctcttga acccggggag aaaccctata aatgccctga gtgtggcaag 60

agtttttcaa ccacaggaaa cttgacagtc caccaacgga cccacaccgg cgagaaacca 120

tacaagtgtc cggagtgtgg taagtctttc tcaagtcctg ccgaccttac cagacatcaa 180

cgcacacata caggtgaaaa accttacaag tgcccagagt gcggaaaaag tttttcacaa 240

tctggcgacc tccgcaggca ccagcgcact cacaccggtg aaaaaccata caagtgtcct 300

gagtgcggga agagttttag tcaacgagct catctggagc gacaccaaag gactcatact 360

ggggagaaac cgtacaaatg tcccgaatgt gggaagagct tctctaccaa gaattccctt 420

acagagcacc agcgcacgca tacgggagag aagccgtata agtgtccgga atgtggcaag 480

agcttttcca gaagtgacca ccttacaacc caccagagga cgcacacc 528

<210> 39
<211> 176
<212> PRT
<213> Artificial Sequence

<220>
<223> ZNF-TRCa Aminoacid

<400> 39
Met Ala Gln Ala Ala Leu Glu Pro Gly Glu Lys Pro Tyr Lys Cys Pro
1 5 10 15
Glu Cys Gly Lys Ser Phe Ser Thr Thr Gly Asn Leu Thr Val His Gln
20 25 30
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
35 40 45
Ser Phe Ser Ser Pro Ala Asp Leu Thr Arg His Gln Arg Thr His Thr
50 55 60
Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln
65 70 75 80
Ser Gly Asp Leu Arg Arg His Gln Arg Thr His Thr Gly Glu Lys Pro
85 90 95
Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Arg Ala His Leu
100 105 110
Glu Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro
115 120 125
Glu Cys Gly Lys Ser Phe Ser Thr Lys Asn Ser Leu Thr Glu His Gln
130 135 140
Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys
145 150 155 160
Ser Phe Ser Arg Ser Asp His Leu Thr Thr His Gln Arg Thr His Thr

165 170 175

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

<220>
<223> AAVS1 site

<400> 40
agacggccgc gtcagagc 18

<210> 41
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc Finger 1

<400> 41
Glu Arg Ser His Leu Arg Glu
1 5

<210> 42
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc Finger 2

<400> 42
Arg Ala Asp Asn Leu Thr Glu
1 5

<210> 43
<211> 7
<212> PRT
<213> Artificial Sequence

<220>

<223> Zinc Finger 3

<400> 43
Ser Arg Arg Thr Cys Arg Ala
1   5

<210> 44
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc Finger 4

<400> 44
Arg Asn Asp Thr Leu Thr Glu
1   5

<210> 45
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc Finger 5

<400> 45
Arg Ser Asp Lys Leu Thr Glu
1   5

<210> 46
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Zinc Finger 6

<400> 46
Gln Leu Ala His Leu Arg Ala
1   5

<210> 47
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Nuclear Localization Signal

<400> 47
atggctccaa agaaaaagag gaaagtggga atccacggag tccccgccgc t 51

<210> 48
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS linker Nucleic Acid

<400> 48
ggtggatctg gcggtggatc tggtggcggt 30

<210> 49
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> GGS linker Amino Acid

<400> 49
Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10

<210> 50
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS4x Linker Nucleic Acid

<400> 50
ggagggagtg gtgggtccgg tggtagtggc ggatcc 36

<210> 51
<211> 12
<212> PRT

<213> Artificial Sequence

<220>
<223> GGS4x Linker Amino Acid

<400> 51
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
1 5 10

<210> 52
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS5x Linker Nucleic Acid

<400> 52
ggaggctccg gtgggtctgg tgggagcggt ggtagtggcg gatcc 45

<210> 53
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> GGS5x Linker Amino Acid

<400> 53
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
1 5 10 15

<210> 54
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS6x Linker Nucleic Acid

<400> 54
ggaggcagtg gtgggagcgg tggttccggg ggtagtggtg gttccggggg atcc 54

<210> 55

<211> 18
<212> PRT
<213> Artificial Sequence

<220>
<223> GGS6x Linker Amino Acid

<400> 55
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1   5   10   15
Gly Ser

<210> 56
<211> 63
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS7x Linker Nucleic Acid

<400> 56
ggaggttctg gaggctccgg tgggtccggg ggaagtgggg ggtcaggcgg atcaggagga 60

tcc 63

<210> 57
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> GGS7x Linker Amino Acid

<400> 57
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1   5   10   15
Gly Ser Gly Gly Ser
            20

<210> 58
<211> 75
<212> DNA
<213> Artificial Sequence

<220>
<223> GGS8x Linker Nucleic Acid

<400> 58
ggaggtagcg gaggttccgg agggagcggc gggagtgggg gaagcggggg aagtggagga 60

tccgggggag gatcc 75

<210> 59
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> GGS8x Linker Amino Acid

<400> 59
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15
Gly Ser Gly Gly Ser
20

<210> 60
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Linker XTEN Nucleic Acid

<400> 60
tccggtagcg aaacaccggg gacttcagaa tcggccaccc cggagtct 48

<210> 61
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker XTEN Amino Acid

<400> 61
Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser
1 5 10 15

<210> 62
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Linker B Nucleic Acid

<400> 62
ggaagcgccg gtagtgcggc tgggtctggc gagttc 36

<210> 63
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker B Amino Acid

<400> 63
Gly Ser Ala Gly Ser Ala Ala Gly Ser Gly Glu Phe
1 5 10

<210> 64
<211> 4104
<212> DNA
<213> Artificial Sequence

<220>
<223> human Cas9 (hCas9) nucleic acid

<400> 64
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gaggggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agac 4104

<210> 65
<211> 4104
<212> DNA
<213> Artificial Sequence

<220>
<223> nickase Cas9 (nCas9) nucleic acid

<400> 65
atggacaaga agtactccat tgggctcgct atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gagggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac aggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agac 4104

<210> 66
<211> 4104
<212> DNA
<213> Artificial Sequence

<220>
<223> dead Cas9 (dCas9) nucleic acid

<400> 66
atggacaaga agtactccat tgggctcgct atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gagggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggctgct 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaagcta gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agac 4104

<210> 67
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> hyperactive PiggyBac (PB) transposase nucleic acid

<400> 67
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 68
<211> 1020
<212> DNA
<213> Artificial Sequence

<220>
<223> hyperactive Sleeping Beauty (SB100) transposase nucleic acid

<400> 68
atgggaaaat caaaagaaat cagccaagac ctcagaaaaa gaattgtaga cctccacaag 60

tctggttcat ccttgggagc aatttccaaa cgcctggcgg taccacgttc atctgtacaa 120

acaatagtac gcaagtataa acaccatggg accacgcagc cgtcataccg ctcaggaagg 180

agacgcgttc tgtctcctag agatgaacgt actttggtgc gaaaagtgca aatcaatccc 240

agaacaacag caaaggacct tgtgaagatg ctggaggaaa caggtacaaa agtatctata 300

tccacagtaa aacgagtcct atatcgacat aacctgaaag gccactcagc aaggaagaag 360

ccactgctcc aaaaccgaca taagaaagcc agactacggt ttgcaactgc acatggggac 420

aaagatcgta ctttttggag aaatgtcctc tggtctgatg aaacaaaaat agaactgttt 480

ggccataatg accatcgtta tgtttggagg aagaaggggg aggcttgcaa gccgaagaac 540

accatcccaa ccgtgaagca cgggggtggc agcatcatgt tgtgggggtg ctttgctgca 600

ggagggactg gtgcacttca caaaatagat ggcatcatgg acgccgtgca gtatgtggat 660

atattgaagc aacatctcaa gacatcagtc aggaagttaa agcttggtcg caaatgggtc 720

ttccaacacg acaatgaccc caagcatact tccaaagttg tggcaaaatg gcttaaggac 780

aacaaagtca aggtattgga gtggccatca caaagccctg acctcaatcc tatagaaaat 840

ttgtgggcag aactgaaaaa gcgtgtgcga gcaaggaggc ctacaaacct gactcagtta 900

caccagctct gtcaggagga atgggccaaa attcacccaa attattgtgg gaagcttgtg 960

gaaggctacc cgaaacgttt gacccaagtt aaacaattta aaggcaatgc taccaaatac 1020

<210> 69
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> human Cas9 (hCas9) amino acid

<400> 69
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1   5   10   15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20   25   30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35   40   45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50   55   60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65   70   75   80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85   90   95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100   105   110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115   120   125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130   135   140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145   150   155   160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165   170   175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180   185   190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195   200   205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210   215   220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225   230   235   240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245   250   255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260   265   270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275   280   285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290   295   300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305   310   315   320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325   330   335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340   345   350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355   360   365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370   375   380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg

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

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

1185 1190 1195 1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
1205 1210 1215
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
1220 1225 1230
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1265 1270 1275 1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
1285 1290 1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
1300 1305 1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
1315 1320 1325
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
1330 1335 1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345 1350 1355 1360
Asp Leu Ser Gln Leu Gly Gly Asp
1365

<210> 70
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> nickase Cas9 (nCas9) amino acid

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

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

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

```
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
        995                 1000                1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
    1010                1015                1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1025                1030                1035                1040
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
                1045                1050                1055
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
            1060                1065                1070
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
        1075                1080                1085
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
    1090                1095                1100
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1105                1110                1115                1120
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
                1125                1130                1135
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
            1140                1145                1150
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
        1155                1160                1165
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
    1170                1175                1180
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1185                1190                1195                1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
                1205                1210                1215
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
            1220                1225                1230
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
        1235                1240                1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
    1250                1255                1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1265                1270                1275                1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
                1285                1290                1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
            1300                1305                1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
        1315                1320                1325
```

```
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
    1330                1335                1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345                1350                1355                1360
Asp Leu Ser Gln Leu Gly Gly Asp
                1365
```

<210> 71
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> dead Cas9 (dCas9) amino acid

<400> 71

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

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

```
            660                 665                 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
        755                 760                 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830
Leu Ser Asp Tyr Asp Val Ala Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945                 950                 955                 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
            980                 985                 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
        995                 1000                1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
    1010                1015                1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1025                1030                1035                1040
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
                1045                1050                1055
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
```

```
            1060                1065                1070
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
        1075                1080                1085
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
    1090                1095                1100
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1105                1110                1115                1120
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
                1125                1130                1135
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
            1140                1145                1150
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
        1155                1160                1165
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
    1170                1175                1180
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1185                1190                1195                1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
                1205                1210                1215
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
            1220                1225                1230
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
        1235                1240                1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
    1250                1255                1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1265                1270                1275                1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
                1285                1290                1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
            1300                1305                1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
        1315                1320                1325
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
    1330                1335                1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345                1350                1355                1360
Asp Leu Ser Gln Leu Gly Gly Asp
                1365
```

<210> 72
<211> 1069
<212> PRT
<213> Staphylococcus aureus

<400> 72

```
Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1               5                   10                  15
Ala Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
```

```
            20                  25                  30
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
        35                  40                  45
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
    50                  55                  60
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
65                  70                  75                  80
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
                85                  90                  95
Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
            100                 105                 110
Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
        115                 120                 125
Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
    130                 135                 140
Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
145                 150                 155                 160
Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
                165                 170                 175
Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
            180                 185                 190
Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
        195                 200                 205
Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
    210                 215                 220
Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
225                 230                 235                 240
Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
                245                 250                 255
Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
            260                 265                 270
Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
        275                 280                 285
Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
    290                 295                 300
Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
305                 310                 315                 320
Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
                325                 330                 335
Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
            340                 345                 350
Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
        355                 360                 365
Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
    370                 375                 380
Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
385                 390                 395                 400
Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
                405                 410                 415
Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
```

```
            420                 425                 430
Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
        435                 440                 445
Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
    450                 455                 460
Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
465                 470                 475                 480
Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
                485                 490                 495
Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
            500                 505                 510
Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
        515                 520                 525
Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
    530                 535                 540
Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
545                 550                 555                 560
Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
                565                 570                 575
Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
            580                 585                 590
Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
        595                 600                 605
Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
    610                 615                 620
Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
625                 630                 635                 640
Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
                645                 650                 655
Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
            660                 665                 670
Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
        675                 680                 685
Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
    690                 695                 700
Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
705                 710                 715                 720
Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
                725                 730                 735
Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
            740                 745                 750
Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
        755                 760                 765
Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
    770                 775                 780
Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
785                 790                 795                 800
Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
                805                 810                 815
Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
```

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

<210> 73
<211> 340
<212> PRT
<213> Artificial Sequence


<220>
<223> hyperactive Sleeping Beauty (SB100) transposase amino acid

<400> 73
Met Gly Lys Ser Lys Glu Ile Ser Gln Asp Leu Arg Lys Arg Ile Val
1               5                   10                  15
Asp Leu His Lys Ser Gly Ser Ser Leu Gly Ala Ile Ser Lys Arg Leu
            20                  25                  30
Ala Val Pro Arg Ser Ser Val Gln Thr Ile Val Arg Lys Tyr Lys His
        35                  40                  45
His Gly Thr Thr Gln Pro Ser Tyr Arg Ser Gly Arg Arg Arg Val Leu
    50                  55                  60
```

```
Ser Pro Arg Asp Glu Arg Thr Leu Val Arg Lys Val Gln Ile Asn Pro
65                  70                  75                  80
Arg Thr Thr Ala Lys Asp Leu Val Lys Met Leu Glu Glu Thr Gly Thr
                85                  90                  95
Lys Val Ser Ile Ser Thr Val Lys Arg Val Leu Tyr Arg His Asn Leu
            100                 105                 110
Lys Gly His Ser Ala Arg Lys Lys Pro Leu Leu Gln Asn Arg His Lys
        115                 120                 125
Lys Ala Arg Leu Arg Phe Ala Thr Ala His Gly Asp Lys Asp Arg Thr
    130                 135                 140
Phe Trp Arg Asn Val Leu Trp Ser Asp Glu Thr Lys Ile Glu Leu Phe
145                 150                 155                 160
Gly His Asn Asp His Arg Tyr Val Trp Arg Lys Lys Gly Glu Ala Cys
                165                 170                 175
Lys Pro Lys Asn Thr Ile Pro Thr Val Lys His Gly Gly Gly Ser Ile
            180                 185                 190
Met Leu Trp Gly Cys Phe Ala Ala Gly Gly Thr Gly Ala Leu His Lys
        195                 200                 205
Ile Asp Gly Ile Met Asp Ala Val Gln Tyr Val Asp Ile Leu Lys Gln
    210                 215                 220
His Leu Lys Thr Ser Val Arg Lys Leu Lys Leu Gly Arg Lys Trp Val
225                 230                 235                 240
Phe Gln His Asp Asn Asp Pro Lys His Thr Ser Lys Val Val Ala Lys
                245                 250                 255
Trp Leu Lys Asp Asn Lys Val Lys Val Leu Glu Trp Pro Ser Gln Ser
            260                 265                 270
Pro Asp Leu Asn Pro Ile Glu Asn Leu Trp Ala Glu Leu Lys Lys Arg
        275                 280                 285
Val Arg Ala Arg Arg Pro Thr Asn Leu Thr Gln Leu His Gln Leu Cys
    290                 295                 300
Gln Glu Glu Trp Ala Lys Ile His Pro Asn Tyr Cys Gly Lys Leu Val
305                 310                 315                 320
Glu Gly Tyr Pro Lys Arg Leu Thr Gln Val Lys Gln Phe Lys Gly Asn
                325                 330                 335
Ala Thr Lys Tyr
            340
```

<210> 74
<211> 1236
<212> PRT
<213> Artificial Sequence

<220>
<223> cpf1

<400> 74

```
Met Ala Pro Lys Lys Lys Arg Lys Val Ser Lys Leu Glu Lys Phe Thr
1               5                   10                  15
Asn Cys Tyr Ser Leu Ser Lys Thr Leu Arg Phe Lys Ala Ile Pro Val
```

```
            20                  25                  30
Gly Lys Thr Gln Glu Asn Ile Asp Asn Lys Arg Leu Leu Val Glu Asp
        35                  40                  45
Glu Lys Arg Ala Glu Asp Tyr Lys Gly Val Lys Lys Leu Leu Asp Arg
    50                  55                  60
Tyr Tyr Leu Ser Phe Ile Asn Asp Val Leu His Ser Ile Lys Leu Lys
65                  70                  75                  80
Asn Leu Asn Asn Tyr Ile Ser Leu Phe Arg Lys Lys Thr Arg Thr Glu
                85                  90                  95
Lys Glu Asn Lys Glu Leu Glu Asn Leu Glu Ile Asn Leu Arg Lys Glu
            100                 105                 110
Ile Ala Lys Ala Phe Lys Gly Asn Glu Gly Tyr Lys Ser Leu Phe Lys
        115                 120                 125
Lys Asp Ile Ile Glu Thr Ile Leu Pro Glu Phe Leu Asp Asp Lys Asp
    130                 135                 140
Glu Ile Ala Leu Val Asn Ser Phe Asn Gly Phe Thr Thr Ala Phe Thr
145                 150                 155                 160
Gly Phe Phe Asp Asn Arg Glu Asn Met Phe Ser Glu Glu Ala Lys Ser
                165                 170                 175
Thr Ser Ile Ala Phe Arg Cys Ile Asn Glu Asn Leu Thr Arg Tyr Ile
            180                 185                 190
Ser Asn Met Asp Ile Phe Glu Lys Val Asp Ala Ile Phe Asp Lys His
        195                 200                 205
Glu Val Gln Glu Ile Lys Glu Lys Ile Leu Asn Ser Asp Tyr Asp Val
    210                 215                 220
Glu Asp Phe Phe Glu Gly Glu Phe Phe Asn Phe Val Leu Thr Gln Glu
225                 230                 235                 240
Gly Ile Asp Val Tyr Asn Ala Ile Ile Gly Gly Phe Val Thr Glu Ser
                245                 250                 255
Gly Glu Lys Ile Lys Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln
            260                 265                 270
Lys Thr Lys Gln Lys Leu Pro Lys Phe Lys Pro Leu Tyr Lys Gln Val
        275                 280                 285
Leu Ser Asp Arg Glu Ser Leu Ser Phe Tyr Gly Glu Gly Tyr Thr Ser
    290                 295                 300
Asp Glu Glu Val Leu Glu Val Phe Arg Asn Thr Leu Asn Lys Asn Ser
305                 310                 315                 320
Glu Ile Phe Ser Ser Ile Lys Lys Leu Glu Lys Leu Phe Lys Asn Phe
                325                 330                 335
Asp Glu Tyr Ser Ser Ala Gly Ile Phe Val Lys Asn Gly Pro Ala Ile
            340                 345                 350
Ser Thr Ile Ser Lys Asp Ile Phe Gly Glu Trp Asn Val Ile Arg Asp
        355                 360                 365
Lys Trp Asn Ala Glu Tyr Asp Asp Ile His Leu Lys Lys Lys Ala Val
    370                 375                 380
Val Thr Glu Lys Tyr Glu Asp Asp Arg Arg Lys Ser Phe Lys Lys Ile
385                 390                 395                 400
Gly Ser Phe Ser Leu Glu Gln Leu Gln Glu Tyr Ala Asp Ala Asp Leu
                405                 410                 415
Ser Val Val Glu Lys Leu Lys Glu Ile Ile Ile Gln Lys Val Asp Glu
```

```
            420                 425                 430
Ile Tyr Lys Val Tyr Gly Ser Ser Glu Lys Leu Phe Asp Ala Asp Phe
        435                 440                 445
Val Leu Glu Lys Ser Leu Lys Lys Asn Asp Ala Val Val Ala Ile Met
    450                 455                 460
Lys Asp Leu Leu Asp Ser Val Lys Ser Phe Glu Asn Tyr Ile Lys Ala
465                 470                 475                 480
Phe Phe Gly Glu Gly Lys Glu Thr Asn Arg Asp Glu Ser Phe Tyr Gly
                485                 490                 495
Asp Phe Val Leu Ala Tyr Asp Ile Leu Leu Lys Val Asp His Ile Tyr
            500                 505                 510
Asp Ala Ile Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Lys Asp Lys
        515                 520                 525
Phe Lys Leu Tyr Phe Gln Asn Pro Gln Phe Met Gly Gly Trp Asp Lys
    530                 535                 540
Asp Lys Glu Thr Asp Tyr Arg Ala Thr Ile Leu Arg Tyr Gly Ser Lys
545                 550                 555                 560
Tyr Tyr Leu Ala Ile Met Asp Lys Lys Tyr Ala Lys Cys Leu Gln Lys
                565                 570                 575
Ile Asp Lys Asp Asp Val Asn Gly Asn Tyr Glu Lys Ile Asn Tyr Lys
            580                 585                 590
Leu Leu Pro Gly Pro Asn Lys Met Leu Pro Lys Val Phe Phe Ser Lys
        595                 600                 605
Lys Trp Met Ala Tyr Tyr Asn Pro Ser Glu Asp Ile Gln Lys Ile Tyr
    610                 615                 620
Lys Asn Gly Thr Phe Lys Lys Gly Asp Met Phe Asn Leu Asn Asp Cys
625                 630                 635                 640
His Lys Leu Ile Asp Phe Phe Lys Asp Ser Ile Ser Arg Tyr Pro Lys
                645                 650                 655
Trp Ser Asn Ala Tyr Asp Phe Asn Phe Ser Glu Thr Glu Lys Tyr Lys
            660                 665                 670
Asp Ile Ala Gly Phe Tyr Arg Glu Val Glu Glu Gln Gly Tyr Lys Val
        675                 680                 685
Ser Phe Glu Ser Ala Ser Lys Lys Glu Val Asp Lys Leu Val Glu Glu
    690                 695                 700
Gly Lys Leu Tyr Met Phe Gln Ile Tyr Asn Lys Asp Phe Ser Asp Lys
705                 710                 715                 720
Ser His Gly Thr Pro Asn Leu His Thr Met Tyr Phe Lys Leu Leu Phe
                725                 730                 735
Asp Glu Asn Asn His Gly Gln Ile Arg Leu Ser Gly Gly Ala Glu Leu
            740                 745                 750
Phe Met Arg Arg Ala Ser Leu Lys Lys Glu Glu Leu Val Val His Pro
        755                 760                 765
Ala Asn Ser Pro Ile Ala Asn Lys Asn Pro Asp Asn Pro Lys Lys Thr
    770                 775                 780
Thr Thr Leu Ser Tyr Asp Val Tyr Lys Asp Lys Arg Phe Ser Glu Asp
785                 790                 795                 800
Gln Tyr Glu Leu His Ile Pro Ile Ala Ile Asn Lys Cys Pro Lys Asn
                805                 810                 815
Ile Phe Lys Ile Asn Thr Glu Val Arg Val Leu Leu Lys His Asp Asp
```

```
            820                 825                 830
Asn Pro Tyr Val Ile Gly Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr
        835                 840                 845
Ile Val Val Val Asp Gly Lys Gly Asn Ile Val Glu Gln Tyr Ser Leu
    850                 855                 860
Asn Glu Ile Ile Asn Asn Phe Asn Gly Ile Arg Ile Lys Thr Asp Tyr
865                 870                 875                 880
His Ser Leu Leu Asp Lys Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln
                885                 890                 895
Asn Trp Thr Ser Ile Glu Asn Ile Lys Glu Leu Lys Ala Gly Tyr Ile
            900                 905                 910
Ser Gln Val Val His Lys Ile Cys Glu Leu Val Glu Lys Tyr Asp Ala
        915                 920                 925
Val Ile Ala Leu Glu Asp Leu Asn Ser Gly Phe Lys Asn Ser Arg Val
    930                 935                 940
Lys Val Glu Lys Gln Val Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp
945                 950                 955                 960
Lys Leu Asn Tyr Met Val Asp Lys Lys Ser Asn Pro Cys Ala Thr Gly
                965                 970                 975
Gly Ala Leu Lys Gly Tyr Gln Ile Thr Asn Lys Phe Glu Ser Phe Lys
            980                 985                 990
Ser Met Ser Thr Gln Asn Gly Phe Ile Phe Tyr Ile Pro Ala Trp Leu
        995                 1000                1005
Thr Ser Lys Ile Asp Pro Ser Thr Gly Phe Val Asn Leu Leu Lys Thr
    1010                1015                1020
Lys Tyr Thr Ser Ile Ala Asp Ser Lys Lys Phe Ile Ser Ser Phe Asp
1025                1030                1035                1040
Arg Ile Met Tyr Val Pro Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp
                1045                1050                1055
Tyr Lys Asn Phe Ser Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys
            1060                1065                1070
Leu Tyr Ser Tyr Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys
        1075                1080                1085
Asn Asn Val Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr Lys
    1090                1095                1100
Glu Leu Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp Ile Arg
1105                1110                1115                1120
Ala Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser Ser Phe Met
                1125                1130                1135
Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser Ile Thr Gly Arg
            1140                1145                1150
Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys Asn Ser Asp Gly Ile
        1155                1160                1165
Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln Glu Asn Ala Ile Leu Pro
    1170                1175                1180
Lys Asn Ala Asp Ala Asn Gly Ala Tyr Asn Ile Ala Arg Lys Val Leu
1185                1190                1195                1200
Trp Ala Ile Gly Gln Phe Lys Lys Ala Glu Asp Glu Lys Leu Asp Lys
                1205                1210                1215
Val Lys Ile Ala Ile Ser Asn Lys Glu Trp Leu Glu Tyr Ala Gln Thr
```

```
           1220                1225                1230
Ser Val Lys His
        1235

<210> 75
<211> 990
<212> PRT
<213> Artificial Sequence


<220>
<223> CasX

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

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

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

<210> 76
<211> 1069
<212> PRT
<213> Staphylococcus aureus

<400> 76

```
Met Ala Pro Lys Lys Lys Arg Lys Val Gly Ile His Gly Val Pro Ala
1               5                   10                  15
Ala Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
            20                  25                  30
```

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

```
Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
        435                 440                 445
Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
    450                 455                 460
Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
465                 470                 475                 480
Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
                485                 490                 495
Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
            500                 505                 510
Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
        515                 520                 525
Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
    530                 535                 540
Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
545                 550                 555                 560
Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
                565                 570                 575
Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
            580                 585                 590
Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
        595                 600                 605
Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
    610                 615                 620
Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
625                 630                 635                 640
Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
                645                 650                 655
Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
            660                 665                 670
Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
        675                 680                 685
Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
    690                 695                 700
Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
705                 710                 715                 720
Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
                725                 730                 735
Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
            740                 745                 750
Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
        755                 760                 765
Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
    770                 775                 780
Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
785                 790                 795                 800
Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
                805                 810                 815
Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
            820                 825                 830
```

```
Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
        835                 840                 845
Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
    850                 855                 860
Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
865                 870                 875                 880
Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
                885                 890                 895
Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
            900                 905                 910
Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
        915                 920                 925
Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
    930                 935                 940
Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
945                 950                 955                 960
Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
                965                 970                 975
Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
            980                 985                 990
Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
        995                 1000                1005
Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
    1010                1015                1020
Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr
1025                1030                1035                1040
Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu
                1045                1050                1055
Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
            1060                1065
```

<210> 77
<211> 3207
<212> DNA
<213> Staphylococcus aureus

<400> 77

```
atggccccaa agaagaagcg gaaggtcggt atccacggag tcccagcagc caagcggaac       60

tacatcctgg gcctggacat cggcatcacc agcgtgggct acggcatcat cgactacgag      120

acacgggacg tgatcgatgc cggcgtgcgg ctgttcaaag aggccaacgt ggaaaacaac      180

gagggcaggc ggagcaagag aggcgccaga aggctgaagc ggcggaggcg gcatagaatc      240

cagagagtga agaagctgct gttcgactac aacctgctga ccgaccacag cgagctgagc      300

ggcatcaacc cctacgaggc cagagtgaag ggcctgagcc agaagctgag cgaggaagag      360
```

ttctctgccg ccctgctgca cctggccaag agaagaggcg tgcacaacgt gaacgaggtg 420

gaagaggaca ccggcaacga gctgtccacc aaagagcaga tcagccggaa cagcaaggcc 480

ctggaagaga aatacgtggc cgaactgcag ctggaacggc tgaagaaaga cggcgaagtg 540

cggggcagca tcaacagatt caagaccagc gactacgtga aagaagccaa acagctgctg 600

aaggtgcaga aggcctacca ccagctggac cagagcttca tcgacaccta catcgacctg 660

ctggaaaccc ggcggaccta ctatgaggga cctggcgagg gcagcccctt cggctggaag 720

gacatcaaag aatggtacga gatgctgatg ggccactgca cctacttccc cgaggaactg 780

cggagcgtga agtacgccta caacgccgac ctgtacaacg ccctgaacga cctgaacaat 840

ctcgtgatca ccagggacga gaacgagaag ctggaatatt acgagaagtt ccagatcatc 900

gagaacgtgt tcaagcagaa gaagaagccc accctgaagc agatcgccaa agaaatcctc 960

gtgaacgaag aggatattaa gggctacaga gtgaccagca ccggcaagcc cgagttcacc 1020

aacctgaagg tgtaccacga catcaaggac attaccgccc ggaaagagat tattgagaac 1080

gccgagctgc tggatcagat tgccaagatc ctgaccatct accagagcag cgaggacatc 1140

caggaagaac tgaccaatct gaactccgag ctgacccagg aagagatcga gcagatctct 1200

aatctgaagg gctataccgg cacccacaac ctgagcctga aggccatcaa cctgatcctg 1260

gacgagctgt ggcacaccaa cgacaaccag atcgctatct tcaaccggct gaagctggtg 1320

cccaagaagg tggacctgtc ccagcagaaa gagatcccca ccaccctggt ggacgacttc 1380

atcctgagcc ccgtcgtgaa gagaagcttc atccagagca tcaaagtgat caacgccatc 1440

atcaagaagt acggcctgcc caacgacatc attatcgagc tggcccgcga gaagaactcc 1500

aaggacgccc agaaaatgat caacgagatg cagaagcgga accggcagac caacgagcgg 1560

atcgaggaaa tcatccggac caccggcaaa gagaacgcca agtacctgat cgagaagatc 1620

aagctgcacg acatgcagga aggcaagtgc ctgtacagcc tggaagccat ccctctggaa 1680

gatctgctga acaacccctt caactatgag gtggaccaca tcatccccag aagcgtgtcc 1740

ttcgacaaca gcttcaacaa caaggtgctc gtgaagcagg aagaaaacag caagaagggc 1800

aaccggaccc cattccagta cctgagcagc agcgacagca agatcagcta cgaaaccttc 1860

aagaagcaca tcctgaatct ggccaagggc aagggcagaa tcagcaagac caagaaagag 1920

tatctgctgg aagaacggga catcaacagg ttctccgtgc agaaagactt catcaaccgg 1980

aacctggtgg ataccagata cgccaccaga ggcctgatga acctgctgcg gagctacttc 2040

agagtgaaca acctggacgt gaaagtgaag tccatcaatg gcggcttcac cagctttctg 2100

cggcggaagt ggaagtttaa gaaagagcgg aacaaggggt acaagcacca cgccgaggac 2160

gccctgatca ttgccaacgc cgatttcatc ttcaaagagt ggaagaaact ggacaaggcc 2220

aaaaaagtga tggaaaacca gatgttcgag gaaaagcagg ccgagagcat gcccgagatc 2280

gaaaccgagc aggagtacaa agagatcttc atcacccccc accagatcaa gcacattaag 2340

gacttcaagg actacaagta cagccaccgg gtggacaaga agcctaatag agagctgatt 2400

aacgacaccc tgtactccac ccggaaggac gacaagggca acaccctgat cgtgaacaat 2460

ctgaacggcc tgtacgacaa ggacaatgac aagctgaaaa agctgatcaa caagagcccc 2520

gaaaagctgc tgatgtacca ccacgacccc cagacctacc agaaactgaa gctgattatg 2580

gaacagtacg gcgacgagaa gaatcccctg tacaagtact acgaggaaac cgggaactac 2640

ctgaccaagt actccaaaaa ggacaacggc cccgtgatca agaagattaa gtattacggc 2700

aacaaactga acgcccatct ggacatcacc gacgactacc ccaacagcag aaacaaggtc 2760

gtgaagctgt ccctgaagcc ctacagattc gacgtgtacc tggacaatgg cgtgtacaag 2820

ttcgtgaccg tgaagaatct ggatgtgatc aaaaaagaaa actactacga agtgaatagc 2880

aagtgctatg aggaagctaa gaagctgaag aagatcagca accaggccga gtttatcgcc 2940

tccttctaca acaacgatct gatcaagatc aacggcgagc tgtatagagt gatcggcgtg 3000

aacaacgacc tgctgaaccg gatcgaagtg aacatgatcg acatcaccta ccgcgagtac 3060

ctggaaaaca tgaacgacaa gaggcccccc aggatcatta agacaatcgc ctccaagacc 3120

cagagcatta agaagtacag cacagacatt ctgggcaacc tgtatgaagt gaaatctaag 3180

aagcaccctc agatcatcaa aaagggc 3207

<210> 78
<211> 3708
<212> DNA

<213> Artificial Sequence

<220>
<223> cpf1 coding sequence

<400> 78
atggccccaa agaagaagcg gaaggtcagc aagctggaga agtttacaaa ctgctactcc 60

ctgtctaaga ccctgaggtt caaggccatc cctgtgggca agacccagga gaacatcgac 120

aataagcggc tgctggtgga ggacgagaag agagccgagg attataaggg cgtgaagaag 180

ctgctggatc gctactatct gtcttttatc aacgacgtgc tgcacagcat caagctgaag 240

aatctgaaca attacatcag cctgttccgg aagaaaacca gaaccgagaa ggagaataag 300

gagctggaga acctggagat caatctgcgg aaggagatcg ccaaggcctt caagggcaac 360

gagggctaca agtccctgtt taagaaggat atcatcgaga caatcctgcc agagttcctg 420

gacgataagg acgagatcgc cctggtgaac agcttcaatg gctttaccac agccttcacc 480

ggcttctttg ataacagaga gaatatgttt tccgaggagg ccaagagcac atccatcgcc 540

ttcaggtgta tcaacgagaa tctgacccgc tacatctcta atatggacat cttcgagaag 600

gtggacgcca tctttgataa gcacgaggtg caggagatca aggagaagat cctgaacagc 660

gactatgatg tggaggattt ctttgagggc gagttcttta actttgtgct gacacaggag 720

ggcatcgacg tgtataacgc catcatcggc ggcttcgtga ccgagagcgg cgagaagatc 780

aagggcctga acgagtacat caacctgtat aatcagaaaa ccaagcagaa gctgcctaag 840

tttaagccac tgtataagca ggtgctgagc gatcgggagt ctctgagctt ctacggcgag 900

ggctatacat ccgatgagga ggtgctggag gtgtttagaa acaccctgaa caagaacagc 960

gagatcttca gctccatcaa gaagctggag aagctgttca agaattttga cgagtactct 1020

agcgccggca tctttgtgaa gaacggcccc gccatcagca caatctccaa ggatatcttc 1080

ggcgagtgga acgtgatccg ggacaagtgg aatgccgagt atgacgatat ccacctgaag 1140

aagaaggccg tggtgaccga gaagtacgag gacgatcgga gaaagtcctt caagaagatc 1200

ggctcctttt ctctggagca gctgcaggag tacgccgacg ccgatctgtc tgtggtggag 1260

aagctgaagg agatcatcat ccagaaggtg gatgagatct acaaggtgta tggctcctct 1320

gagaagctgt tcgacgccga ttttgtgctg gagaagagcc tgaagaagaa cgacgccgtg 1380

gtggccatca tgaaggacct gctggattct gtgaagagct tcgagaatta catcaaggcc 1440

ttctttggcg agggcaagga gacaaacagg gacgagtcct tctatggcga ttttgtgctg 1500

gcctacgaca tcctgctgaa ggtggaccac atctacgatg ccatccgcaa ttatgtgacc 1560

cagaagccct actctaagga taagttcaag ctgtattttc agaaccctca gttcatgggc 1620

ggctgggaca aggataagga gacagactat cgggccacca tcctgagata cggctccaag 1680

tactatctgg ccatcatgga taagaagtac gccaagtgcc tgcagaagat cgacaaggac 1740

gatgtgaacg gcaattacga gaagatcaac tataagctgc tgcccggccc taataagatg 1800

ctgccaaagg tgttcttttc taagaagtgg atggcctact ataaccccag cgaggacatc 1860

cagaagatct acaagaatgg cacattcaag aagggcgata tgtttaacct gaatgactgt 1920

cacaagctga tcgacttctt taaggatagc atctcccggt atccaaagtg gtccaatgcc 1980

tacgatttca acttttctga gacagagaag tataaggaca tcgccggctt ttacagagag 2040

gtggaggagc agggctataa ggtgagcttc gagtctgcca gcaagaagga ggtggataag 2100

ctggtggagg agggcaagct gtatatgttc cagatctata acaaggactt ttccgataag 2160

tctcacggca cacccaatct gcacaccatg tacttcaagc tgctgtttga cgagaacaat 2220

cacggacaga tcaggctgag cggaggagca gagctgttca tgaggcgcgc ctccctgaag 2280

aaggaggagc tggtggtgca cccagccaac tcccctatcg ccaacaagaa tccagataat 2340

cccaagaaaa ccacaaccct gtcctacgac gtgtataagg ataagaggtt ttctgaggac 2400

cagtacgagc tgcacatccc aatcgccatc aataagtgcc ccaagaacat cttcaagatc 2460

aatacagagg tgcgcgtgct gctgaagcac gacgataacc cctatgtgat cggcatcgat 2520

aggggcgagc gcaatctgct gtatatcgtg gtggtggacg gcaagggcaa catcgtggag 2580

cagtattccc tgaacgagat catcaacaac ttcaacggca tcaggatcaa gacagattac 2640

cactctctgc tggacaagaa ggagaaggag aggttcgagg cccgccagaa ctggacctcc 2700

atcgagaata tcaaggagct gaaggccggc tatatctctc aggtggtgca caagatctgc 2760

gagctggtgg agaagtacga tgccgtgatc gccctggagg acctgaactc tggctttaag 2820

aatagccgcg tgaaggtgga gaagcaggtg tatcagaagt tcgagaagat gctgatcgat 2880

aagctgaact acatggtgga caagaagtct aatccttgtg caacaggcgg cgccctgaag 2940

ggctatcaga tcaccaataa gttcgagagc tttaagtcca tgtctaccca gaacggcttc 3000

atcttttaca tccctgcctg gctgacatcc aagatcgatc catctaccgg ctttgtgaac 3060

ctgctgaaaa ccaagtatac cagcatcgcc gattccaaga agttcatcag ctcctttgac 3120

aggatcatgt acgtgcccga ggaggatctg ttcgagtttg ccctggacta taagaacttc 3180

tctcgcacag acgccgatta catcaagaag tggaagctgt actcctacgg caaccggatc 3240

agaatcttcc ggaatcctaa gaagaacaac gtgttcgact gggaggaggt gtgcctgacc 3300

agcgcctata aggagctgtt caacaagtac ggcatcaatt atcagcaggg cgatatcaga 3360

gccctgctgt gcgagcagtc cgacaaggcc ttctactcta gctttatggc cctgatgagc 3420

ctgatgctgc agatgcggaa cagcatcaca ggccgcaccg acgtggattt tctgatcagc 3480

cctgtgaaga actccgacgg catcttctac gatagccgga actatgaggc ccaggagaat 3540

gccatcctgc caaagaacgc cgacgccaat ggcgcctata acatcgccag aaaggtgctg 3600

tgggccatcg gccagttcaa gaaggccgag gacgagaagc tggataaggt gaagatcgcc 3660

atctctaaca aggagtggct ggagtacgcc cagaccagcg tgaagcac 3708

<210> 79
<211> 2970
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX coding sequence

<400> 79
atggccccaa agaagaagcg gaaggtcagc atgcaagaga tcaagagaat caacaagatc 60

agaaggagac tggtcaagga cagcaacaca aagaaggccg gcaagacagg ccccatgaaa 120

accctgctcg tcagagtgat gacccctgac ctgagagagc ggctggaaaa cctgagaaag 180

aagcccgaga acatccctca gcctatcagc aacaccagca gggccaacct gaacaagctg 240

ctgaccgact acaccgagat gaagaaagcc atcctgcacg tgtactggga agagttccag 300

aaagaccccg tgggcctgat gagcagagtt gctcagcccg ctcctaagaa catcgaccag 360

agaaagctga tccccgtgaa ggacggcaac gagagactga cctctagcgg ctttgcctgc 420

agccagtgtt gccagcctct gtacgtgtac aagctggaac aagtgaacga caagggcaag 480

ccccacacca actacttcgg cagatgcaac gtgtccgagc acgagaggct gatcctgctg 540

tctcctcaca agcccgaggc caacgatgag ctggtcacat acagcctggg caagttcgga 600

cagagagccc tggacttcta cagcatccac gtgaccaggg agagcaatca ccctgtgaag 660

cccctggaac agatcggcgg caatagctgt gcctctggac ctgtgggaaa agccctgagc 720

gacgcctgta tgggagccgt ggcatccttc ctgaccaagt accaggacat catcctggaa 780

caccagaaag tgatcaagaa gaacgagaaa agactggcca acctcaagga tatcgccagc 840

gctaacggcc tggcctttcc taagatcacc ctgcctccac agcctcacac caaagagggc 900

atcgaggcct acaacaacgt ggtggcccag atcgtgattt gggtcaacct gaatctgtgg 960

cagaagctga agatcggcag ggacgaagcc aagccactgc agagactgaa gggcttccct 1020

agcttccctc tggtggaaag acaggccaat gaagtggatt ggtgggacat ggtctgcaac 1080

gtgaagaagc tgatcaacga gaagaaagag gatggcaagg ttttctggca gaacctggcc 1140

ggctacaaga gacaagaagc cctgctgcct tacctgagca gcgaagagga ccggaagaag 1200

ggcaagaagt tcgccagata ccagttcggc gacctgctgc tgcacctgga aaagaagcac 1260

ggcgaggact ggggcaaagt gtacgatgag gcctgggaga gaatcgacaa gaaggtggaa 1320

ggcctgagca agcacattaa gctggaagag gaaagaagga gcgaggacgc ccaatctaaa 1380

gccgctctga ccgattggct gagagccaag gccagctttg tgatcgaggg cctgaaagag 1440

gccgacaagg acgagttctg cagatgcgag ctgaagctgc agaagtggta cggcgatctg 1500

agaggcaagc ccttcgccat tgaggccgag aacagcatcc tggacatcag cggcttcagc 1560

aagcagtaca actgcgcctt catttggcag aaagacggcg tcaagaaact gaacctgtac 1620

ctgatcatca attacttcaa aggcggcaag ctgcggttca agaagatcaa acccgaggcc 1680

ttcgaggcta acagattcta caccgtgatc aacaaaaagt ccggcgagat cgtgcccatg 1740

gaagtgaact tcaacttcga cgaccccaac ctgattatcc tgcctctggc cttcggcaag 1800

agacagggca gagagttcat ctggaacgat ctgctgagcc tggaaaccgg ctctctgaag 1860

ctggccaatg gcagagtgat cgagaaaacc ctgtacaaca ggagaaccag acaggacgag 1920

cctgctctgt ttgtggccct gaccttcgag agaagagagg tgctggacag cagcaacatc 1980

aagcccatga acctgatcgg catcgaccgg ggcgagaata tccctgctgt gatcgccctg 2040

acagaccctg aaggatgccc actgagcaga ttcaaggact ccctgggcaa ccctacacac 2100

atcctgagaa tcggcgagag ctacaaagag aagcagagga caatccaggc cgccaaagag 2160

gtggaacaga gaagagccgg cggatactct aggaagtacg ccagcaaggc caagaatctg 2220

gccgacgaca tggtccgaaa caccgccaga gatctgctgt actacgccgt gacacaggac 2280

gccatgctga tcttcgagaa tctgagcaga ggcttcggcc ggcagggcaa gagaaccttt 2340

atggccgaga ggcagtacac cagaatggaa gattggctca cagctaaact ggcctacgag 2400

ggactgccca gcaagaccta cctgtccaaa acactggccc agtatacctc caagacctgc 2460

agcaattgcg gcttcaccat caccagcgcc gactacgaca gagtgctgga aaagctcaag 2520

aaaaccgcca ccggctggat gaccaccatc aacggcaaag agctgaaggt tgagggccag 2580

atcacctact acaacaggta caagaggcag aacgtcgtga aggatctgag cgtggaactg 2640

gacagactga gcgaagagag cgtgaacaac gacatcagca gctggacaaa gggcagatca 2700

ggcgaggctc tgagcctgct gaagaagagg tttagccaca gacctgtgca agagaagttc 2760

gtgtgcctga actgcggctt cgagacacac gccgatgaac aggctgccct gaacattgcc 2820

agaagctggc tgttcctgag aagccaagag tacaagaagt accagaccaa caagaccacc 2880

ggcaacaccg acaagagggc ctttgtggaa acctggcaga gcttctacag aaaaaagctg 2940

aaagaagtct ggaagcccgc cgtgactagt 2970

<210> 80
<211> 3213
<212> DNA
<213> Staphylococcus aureus

<400> 80

gccaccatgg ccccaaagaa gaagcggaag gtcggtatcc acggagtccc agcagccaag 60

cggaactaca tcctgggcct ggccatcggc atcaccagcg tgggctacgg catcatcgac 120

tacgagacac gggacgtgat cgatgccggc gtgcggctgt tcaaagaggc caacgtggaa 180

aacaacgagg gcaggcggag caagagaggc gccagaaggc tgaagcggcg gaggcggcat 240

agaatccaga gagtgaagaa gctgctgttc gactacaacc tgctgaccga ccacagcgag 300

ctgagcggca tcaaccccta cgaggccaga gtgaagggcc tgagccagaa gctgagcgag 360

gaagagttct ctgccgccct gctgcacctg gccaagagaa gaggcgtgca caacgtgaac 420

gaggtggaag aggacaccgg caacgagctg tccaccaaag agcagatcag ccggaacagc 480

aaggccctgg aagagaaata cgtggccgaa ctgcagctgg aacggctgaa gaaagacggc 540

gaagtgcggg gcagcatcaa cagattcaag accagcgact acgtgaaaga agccaaacag 600

ctgctgaagg tgcagaaggc ctaccaccag ctggaccaga gcttcatcga cacctacatc 660

gacctgctgg aaacccggcg gacctactat gagggacctg gcgagggcag ccccttcggc 720

tggaaggaca tcaaagaatg gtacgagatg ctgatgggcc actgcaccta cttccccgag 780

gaactgcgga gcgtgaagta cgcctacaac gccgacctgt acaacgccct gaacgacctg 840

aacaatctcg tgatcaccag ggacgagaac gagaagctgg aatattacga gaagttccag 900

atcatcgaga acgtgttcaa gcagaagaag aagcccaccc tgaagcagat cgccaaagaa 960

atcctcgtga acgaagagga tattaagggc tacagagtga ccagcaccgg caagcccgag 1020

ttcaccaacc tgaaggtgta ccacgacatc aaggacatta ccgcccggaa agagattatt 1080

gagaacgccg agctgctgga tcagattgcc aagatcctga ccatctacca gagcagcgag 1140

gacatccagg aagaactgac caatctgaac tccgagctga cccaggaaga gatcgagcag 1200

atctctaatc tgaagggcta taccggcacc cacaacctga gcctgaaggc catcaacctg 1260

atcctggacg agctgtggca caccaacgac aaccagatcg ctatcttcaa ccggctgaag 1320

ctggtgccca agaaggtgga cctgtcccag cagaaagaga tccccaccac cctggtggac 1380

gacttcatcc tgagccccgt cgtgaagaga agcttcatcc agagcatcaa agtgatcaac 1440

gccatcatca agaagtacgg cctgcccaac gacatcatta tcgagctggc ccgcgagaag 1500

aactccaagg acgcccagaa aatgatcaac gagatgcaga agcggaaccg gcagaccaac 1560

gagcggatcg aggaaatcat ccggaccacc ggcaaagaga acgccaagta cctgatcgag 1620

aagatcaagc tgcacgacat gcaggaaggc aagtgcctgt acagcctgga agccatccct 1680

ctggaagatc tgctgaacaa ccccttcaac tatgaggtgg accacatcat ccccagaagc 1740

gtgtccttcg acaacagctt caacaacaag gtgctcgtga agcaggaaga aaacagcaag 1800

aagggcaacc ggaccccatt ccagtacctg agcagcagcg acagcaagat cagctacgaa 1860

accttcaaga agcacatcct gaatctggcc aagggcaagg gcagaatcag caagaccaag 1920

aaagagtatc tgctggaaga acgggacatc aacaggttct ccgtgcagaa agacttcatc 1980

aaccggaacc tggtggatac cagatacgcc accagaggcc tgatgaacct gctgcggagc 2040

tacttcagag tgaacaacct ggacgtgaaa gtgaagtcca tcaatggcgg cttcaccagc 2100

tttctgcggc ggaagtggaa gtttaagaaa gagcggaaca aggggtacaa gcaccacgcc 2160

gaggacgccc tgatcattgc caacgccgat ttcatcttca aagagtggaa gaaactggac 2220

aaggccaaaa aagtgatgga aaaccagatg ttcgaggaaa agcaggccga gagcatgccc 2280

gagatcgaaa ccgagcagga gtacaaagag atcttcatca ccccccacca gatcaagcac 2340

attaaggact tcaaggacta caagtacagc caccgggtgg acaagaagcc taatagagag 2400

ctgattaacg acaccctgta ctccacccgg aaggacgaca agggcaacac cctgatcgtg 2460

aacaatctga acggcctgta cgacaaggac aatgacaagc tgaaaaagct gatcaacaag 2520

agccccgaaa agctgctgat gtaccaccac gacccccaga cctaccagaa actgaagctg 2580

attatggaac agtacggcga cgagaagaat cccctgtaca agtactacga ggaaaccggg 2640

aactacctga ccaagtactc caaaaaggac aacggccccg tgatcaagaa gattaagtat 2700

tacggcaaca aactgaacgc ccatctggac atcaccgacg actaccccaa cagcagaaac 2760

aaggtcgtga agctgtccct gaagccctac agattcgacg tgtacctgga caatggcgtg 2820

tacaagttcg tgaccgtgaa gaatctggat gtgatcaaaa aagaaaacta ctacgaagtg 2880

aatagcaagt gctatgagga agctaagaag ctgaagaaga tcagcaacca ggccgagttt 2940

atcgcctcct tctacaacaa cgatctgatc aagatcaacg gcgagctgta tagagtgatc 3000

ggcgtgaaca acgacctgct gaaccggatc gaagtgaaca tgatcgacat cacctaccgc 3060

gagtacctgg aaaacatgaa cgacaagagg ccccccagga tcattaagac aatcgcctcc 3120

aagacccaga gcattaagaa gtacagcaca gacattctgg gcaacctgta tgaagtgaaa 3180

tctaagaagc accctcagat catcaaaaag ggc 3213

<210> 81
<211> 2976
<212> DNA
<213> Campylobacter jejuni

<400> 81
atggccccaa agaagaagcg gaaggtcgcc agaatcctgg ccttcgacat cggcatcagc 60

agcatcggct gggccttcag cgagaacgac gagctgaagg actgcggcgt gcggatcttc 120

accaaggtgg aaaaccccaa gaccggcgag agcctggccc tgcccagaag gctggccaga 180

agcgcccgga agagactggc cagacggaag gcccggctga accacctgaa gcacctgatc 240

gccaacgagt tcaagctgaa ctacgaggac taccagagct tcgacgagtc cctggccaag 300

gcctacaagg gcagcctgat cagcccctac gagctgcggt tccgggccct gaacgagctg 360

ctgagcaagc aggacttcgc cagagtgatc ctgcacattg ccaagcggag aggctacgac 420

gacatcaaga acagcgacga caaagagaag ggcgccatcc tgaaggccat caagcagaac 480

gaggaaaagc tggccaacta ccagtccgtg ggcgagtacc tgtacaaaga gtacttccag 540

aagttcaaag agaacagcaa agaattcacc aacgtgcgga acaagaaaga aagctacgag 600

cggtgtatcg cccagagctt cctgaaggat gagctgaagc tgatcttcaa gaagcagaga 660

gagttcggct tcagcttcag caagaaattc gaggaagagg tgctgagcgt cgccttctac 720

aagagagccc tgaaggactt cagccacctc gtgggcaact gcagcttctt caccgacgag 780

aagagagccc ccaagaacag ccccctggcc ttcatgttcg tggccctgac ccggatcatc 840

aacctgctga acaatctgaa gaacaccgag ggcatcctgt acaccaagga cgacctgaac 900

gccctgctga atgaggtgct gaagaacggc accctgacct acaagcagac caagaagctg 960

ctgggcctga gcgacgacta cgagtttaag ggcgagaagg gcacctactt catcgagttc 1020

aagaagtaca aagagttcat caaggccctg ggcgagcaca acctgagcca ggacgatctg 1080

aatgagatcg ccaaggacat caccctgatc aaggacgaga ttaagctgaa gaaggccctg 1140

gccaaatacg acctgaatca gaaccagatc gacagcctga gcaagctgga attcaaggat 1200

cacctgaaca tcagcttcaa ggctctgaag ctggtcaccc ccctgatgct ggaaggcaag 1260

aagtacgacg aggcctgcaa cgagctgaac ctgaaggtgg ccatcaacga ggacaagaag 1320

gacttcctgc ccgccttcaa cgaaacctac tacaaggacg aagtgaccaa ccccgtggtg 1380

ctgcgggcca tcaaagaata ccggaaggtg ctgaatgccc tgctcaagaa atacggcaag 1440

gtgcacaaga tcaacatcga gctggcccgg gaagtgggca agaaccacag ccagcgggcc 1500

aagatcgaga aagagcagaa cgaaaactac aaggccaaga aggacgctga gctggaatgc 1560

gagaagctgg gactgaagat caacagcaag aacatcctga agctgcggct gttcaaagaa 1620

cagaaagagt tctgcgccta cagcggcgag aagatcaaga tcagcgatct gcaggacgag 1680

aagatgctgg aaatcgacca catctacccc tacagccggt ccttcgacga cagctacatg 1740

aacaaggtgc tggtgttcac caaacagaac caggaaaaac tgaaccagac ccccttcgag 1800

gccttcggca acgacagcgc caagtggcag aaaatcgagg tgctggccaa gaacctgccc 1860

accaagaaac agaagagaat cctggacaag aattacaagg acaaagagca gaagaacttc 1920

aaggaccgga acctgaacga cacccggtat atcgcccggc tggtgctgaa ctacacaaag 1980

gactacctgg atttcctgcc cctgtccgac gacgagaaca ccaagctgaa cgatacccag 2040

aaaggctcca aggtgcacgt ggaagccaag agcggcatgc tgaccagcgc cctgagacac 2100

acctggggct tcagcgccaa ggatcggaac aaccatctgc accacgccat cgacgccgtg 2160

atcattgcct acgccaacaa cagcatcgtg aaggccttct ccgacttcaa gaaagaacag 2220

gaaagcaaca gcgccgagct gtacgccaag aagatctctg agctggacta caagaacaag 2280

cggaagttct tcgagccctt cagcggcttc cggcagaagg tgctggataa gatcgacgag 2340

atcttcgtgt ccaagcccga gcggaagaag ccctctggcg ccctgcacga ggaaaccttc 2400

agaaaagagg aagagttcta ccagtcctac ggcggcaaag aaggcgtgct gaaggccctc 2460

gagctgggca agatcagaaa agtgaacggc aagatcgtga agaacgggga catgttccgg 2520

gtggacatct tcaagcacaa aaagaccaac aagttctacg ccgtgcccat ctacaccatg 2580

gacttcgccc tgaaggtgct gcccaacaag gccgtggccc ggtccaagaa gggcgagatc 2640

aaggactgga ttctgatgga cgagaactac gagttctgct ttagcctgta caaggactcc 2700

ctgatcctga tccagaccaa ggacatgcag gaacccgagt tcgtctacta caacgccttc 2760

accagcagca ccgtgtccct gatcgtgtct aagcacgaca acaagttcga gacactgagc 2820

aagaaccaga agatcctgtt caagaacgcc aacgagaaag aagtgatcgc caagagcatc 2880

ggcatccaga atctgaaggt gttcgagaag tacatcgtgt ccgccctggg agaagtgaca 2940

aaggccgagt tccggcagag agaggacttc aaaaag 2976

<210> 82
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac D450N/R372A/K375A

<400> 82
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtggccagca acgccagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctgaac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 83
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R275A/R277A

<400> 83
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcgccggcgc ctgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 84
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac K409A/K412A

<400> 84
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgcccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccgccccc gccgccatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 85
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R460A/K461A

<400> 85
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcgcc 1380

gccaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 86
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac S351A/R372A/K375A/R388A/
D450N/W465A/S573A/M589V/S594L

<400> 86
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc gccatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtggccagca acgccagaga gatccccgag 1140

gtcctgaaga acagcaggtc cgcccccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctgaac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca gggcccccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccgcct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacgtgtgc cagaggtgtc tc 1782

<210> 87
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac
R275A/R277A/N347S/K375A/T560A/S573A/M589V/S594L

<400> 87
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcgccggcgc ctgcccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc cgcccccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctgaac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca gggcccccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccgcct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacgtgtgc cagagctgtc tc 1782

<210> 88
<211> 1133
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered PiggyBac R245A/R275A/R277A/R372A/W465A

<400> 88
tcgacttcct gatcagatgc ctgaggatgg acgacaagag catcaggccc accctgcggg 60

agaacgacgt gttcaccccc gtggccaaga tctgggacct gttcatccac cagtgcatcc 120

agaactacac ccctggcgcc cacctgacca tcgacgagca gctgctgggc ttcgccggcg 180

cctgcccctt cagggtctat atccccaaca agcccagcaa gtacggcatc aagatcctga 240

tgatgtgcga cagcggcacc aagtacatga tcaacggcat gccctacctg ggcaggggca 300

cccagaccaa cggcgtgccc ctgggcgagt actacgtgaa ggagctgtcc aagcccgtcc 360

acggcagctg cagaaacatc acctgcgaca actggttcac cagcatcccc ctggccaaga 420

acctgctgca ggagccctac aagctgacca tcgtgggcac cgtggccagc aacaagagag 480

agatccccga ggtcctgaag aacagcaggt ccaggcccgt gggcaccagc atgttctgct 540

tcgacggccc cctgaccctg gtgtcctaca agcccaagcc cgccaagatg gtgtacctgc 600

tgtccagctg cgacgaggac gccagcatca acgagagcac cggcaagccc cagatggtga 660

tgtactacaa ccagaccaag ggcggcgtgg acaccctgga ccagatgtgc agcgtgatga 720

cctgcagcag aaagaccaac agggccccca tggccctgct gtacggcatg atcaacatcg 780

cctgcatcaa cagcttcatc atctacagcc acaacgtgag cagcaagggc gagaaggtgc 840

agagccggaa aaagttcatg cggaacctgt acatgggcct gacctccagc ttcatgagga 900

agaggctgga ggcccccacc ctgaagagat acctgaggga caacatcagc aacatcctgc 960

ccaaagaggt gcccggcacc agcgacgaca gcaccgagga gcccgtgatg aagaagagga 1020

cctactgcac ctactgtccc agcaagatca gaagaaaggc cagcgccagc tgcaagaagt 1080

gtaagaaggt catctgccgg gagcacaaca tcgacatgtg ccagagctgt ttc 1133

<210> 89
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered Piggybac
R372A/K375A/R388A/D450N/W465A/S573A/M589V/S594L

<400> 89
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtggccagca acgccagaga gatccccgag 1140

gtcctgaaga acagcaggtc cgcccccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctgaac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca gggcccccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccgcct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacgtgtgc cagaggtgtc tc 1782

<210> 90
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac M194A

<400> 90
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190

```
His Ala Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Arg Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
```

Cys Phe

<210> 91
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R245A

<400> 91
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Ala Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
245 250 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
260 265 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
275 280 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys

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

<210> 92
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R325A

<400> 92

```
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1               5                  10                  15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50                  55                  60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65                  70                  75                  80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85                  90                  95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100                 105                 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115                 120                 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130                 135                 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145                 150                 155                 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165                 170                 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Ala Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Arg Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
```

Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
405 410 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
420 425 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
435 440 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
450 455 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465 470 475 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
485 490 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
500 505 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
515 520 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
530 535 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545 550 555 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
565 570 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
580 585 590
Cys Phe

<210> 93
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R372A

<400> 93
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu

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

```
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 94
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R375A

<400> 94

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

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

<210> 95
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R376A

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

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

<210> 96
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac E377A

<400> 96
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15

```
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50                  55                  60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65                  70                  75                  80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85                  90                  95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100                 105                 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115                 120                 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130                 135                 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145                 150                 155                 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165                 170                 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180                 185                 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195                 200                 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210                 215                 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225                 230                 235                 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245                 250                 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260                 265                 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275                 280                 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290                 295                 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305                 310                 315                 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325                 330                 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
            340                 345                 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355                 360                 365
Gly Thr Val Arg Ser Asn Lys Arg Ala Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
```

```
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asp Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 97
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac E380A

<400> 97

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

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

```
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 98
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac D450N

<400> 98

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

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

<210> 99

<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Engineered Piggybac S564P

<400> 99
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305   310   315   320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val

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

<210> 100
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered Piggybac

<400> 100
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccacg ccagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 101
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered Piggybac

<400> 101
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tggccaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 102
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered Piggybac

<400> 102
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgcccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tggccgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 103

<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> engineered Piggybac

<400> 103
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtggccagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 104
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R375A

<400> 104
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgcccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acgccagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 105
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R376A

<400> 105
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaaggccga tccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 106
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac E377A

<400> 106
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagagc catccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 107
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac E380A

<400> 107
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgcccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgcc 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 108
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac D450N

<400> 108
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctgaac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtccca gcaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 109
<211> 1782
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac S564P

<400> 109
atgggcagca gcctggacga cgagcacatc ctgagcgccc tgctgcagag cgacgacgag 60

ctggtcggcg aggacagcga cagcgaggtg agcgaccacg tgagcgagga cgacgtgcag 120

tccgacaccg aggaggcctt catcgacgag gtgcacgagg tgcagcctac cagcagcggc 180

tccgagatcc tggacgagca gaacgtgatc gagcagcccg gcagctccct ggccagcaac 240

aggatcctga ccctgcccca gaggaccatc aggggcaaga acaagcactg ctggtccacc 300

tccaagccca ccaggcggag cagggtgtcc gccctgaaca tcgtgagaag ccagaggggc 360

cccaccagga tgtgcaggaa catctacgac cccctgctgt gcttcaagct gttcttcacc 420

gacgagatca tcagcgagat cgtgaagtgg accaacgccg agatcagcct gaagaggcgg 480

gagagcatga cctccgccac cttcagggac accaacgagg acgagatcta cgccttcttc 540

ggcatcctgg tgatgaccgc cgtgaggaag gacaaccaca tgagcaccga cgacctgttc 600

gacagatccc tgagcatggt gtacgtgagc gtgatgagca gggacagatt cgacttcctg 660

atcagatgcc tgaggatgga cgacaagagc atcaggccca ccctgcggga gaacgacgtg 720

ttcacccccg tgagaaagat ctgggacctg ttcatccacc agtgcatcca gaactacacc 780

cctggcgccc acctgaccat cgacgagcag ctgctgggct tcaggggcag gtgccccttc 840

agggtctata tccccaacaa gcccagcaag tacggcatca agatcctgat gatgtgcgac 900

agcggcacca agtacatgat caacggcatg ccctacctgg gcaggggcac ccagaccaac 960

ggcgtgcccc tgggcgagta ctacgtgaag gagctgtcca agcccgtcca cggcagctgc 1020

agaaacatca cctgcgacaa ctggttcacc agcatccccc tggccaagaa cctgctgcag 1080

gagccctaca agctgaccat cgtgggcacc gtgagaagca acaagagaga gatccccgag 1140

gtcctgaaga acagcaggtc caggcccgtg ggcaccagca tgttctgctt cgacggcccc 1200

ctgaccctgg tgtcctacaa gcccaagccc gccaagatgg tgtacctgct gtccagctgc 1260

gacgaggacg ccagcatcaa cgagagcacc ggcaagcccc agatggtgat gtactacaac 1320

cagaccaagg gcggcgtgga caccctggac cagatgtgca gcgtgatgac ctgcagcaga 1380

aagaccaaca ggtggcccat ggccctgctg tacggcatga tcaacatcgc ctgcatcaac 1440

agcttcatca tctacagcca caacgtgagc agcaagggcg agaaggtgca gagccggaaa 1500

aagttcatgc ggaacctgta catgggcctg acctccagct tcatgaggaa gaggctggag 1560

gcccccaccc tgaagagata cctgagggac aacatcagca acatcctgcc caaagaggtg 1620

cccggcacca gcgacgacag caccgaggag cccgtgatga agaagaggac ctactgcacc 1680

tactgtcccc ccaagatcag aagaaaggcc agcgccagct gcaagaagtg taagaaggtc 1740

atctgccggg agcacaacat cgacatgtgc cagagctgtt tc 1782

<210> 110
<211> 5931
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac
S351A/R372A/K375A/R388A/D450N/W465A/S573A/M589V/592G/S594L

<400> 110
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga aacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gagggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agattccggt agcgaaacac cggggacttc agaatcggcc 4140

accccggagt ctggcagcag cctggacgac gagcacatcc tgagcgccct gctgcagagc 4200

gacgacgagc tggtcggcga ggacagcgac agcgaggtga gcgaccacgt gagcgaggac 4260

gacgtgcagt ccgacaccga ggaggccttc atcgacgagg tgcacgaggt gcagcctacc 4320

agcagcggct ccgagatcct ggacgagcag aacgtgatcg agcagcccgg cagctccctg 4380

gccagcaaca ggatcctgac cctgccccag aggaccatca ggggcaagaa caagcactgc 4440

tggtccacct ccaagcccac caggcggagc agggtgtccg ccctgaacat cgtgagaagc 4500

cagaggggcc ccaccaggat gtgcaggaac atctacgacc ccctgctgtg cttcaagctg 4560

ttcttcaccg acgagatcat cagcgagatc gtgaagtgga ccaacgccga gatcagcctg 4620

aagaggcggg agagcatgac ctccgccacc ttcagggaca ccaacgagga cgagatctac 4680

gccttcttcg gcatcctggt gatgaccgcc gtgaggaagg acaaccacat gagcaccgac 4740

gacctgttcg acagatccct gagcatggtg tacgtgagcg tgatgagcag ggacagattc 4800

gacttcctga tcagatgcct gaggatggac gacaagagca tcaggcccac cctgcgggag 4860

aacgacgtgt tcacccccgt gagaaagatc tgggacctgt tcatccacca gtgcatccag 4920

aactacaccc ctggcgccca cctgaccatc gacgagcagc tgctgggctt caggggcagg 4980

tgccccttca gggtctatat ccccaacaag cccagcaagt acggcatcaa gatcctgatg 5040

atgtgcgaca gcggcaccaa gtacatgatc aacggcatgc cctacctggg caggggcacc 5100

cagaccaacg gcgtgccccc gggcgagtac tacgtgaagg agctgtccaa gcccgtccac 5160

ggcagctgca gaaacatcac ctgcgacaac tggttcaccg ccatcccccc ggccaagaac 5220

ctgctgcagg agccctacaa gctgaccatc gtgggcaccg tggccagcaa cgccagagag 5280

atccccgagg tcctgaagaa cagcaggtcc gcccccgtgg gcaccagcat gttctgcttc 5340

gacggccccc tgaccctggt gtcctacaag cccaagcccg ccaagatggt gtacctgctg 5400

tccagctgcg acgaggacgc cagcatcaac gagagcaccg gcaagcccca gatggtgatg 5460

tactacaacc agaccaaggg cggcgtggac accctgaacc agatgtgcag cgtgatgacc 5520

tgcagcagaa agaccaacag ggcccccatg gccctgctgt acggcatgat caacatcgcc 5580

tgcatcaaca gcttcatcat ctacagccac aacgtgagca gcaagggcga gaaggtgcag 5640

agccggaaaa agttcatgcg gaacctgtac atgggcctga cctccagctt catgaggaag 5700

aggctggagg cccccaccct gaagagatac ctgagggaca acatcagcaa catcctgccc 5760

aaagaggtgc ccggcaccag cgacgacagc accgaggagc ccgtgatgaa gaagaggacc 5820

tactgcacct actgtcccag caagatcaga agaaaggcca gcgccgcctg caagaagtgt 5880

aagaaggtca tctgccggga gcacaacatc gacgtgtgcc agggctgttt g 5931

<210> 111
<211> 5931
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R245A/R275A/R277A/R372A/W465A/M589V

<400> 111
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga aacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gagggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agattccggt agcgaaacac cggggacttc agaatcggcc 4140

accccggagt ctggcagcag cctggacgac gagcacatcc tgagcgccct gctgcagagc 4200

gacgacgagc tggtcggcga ggacagcgac agcgaggtga gcgaccacgt gagcgaggac 4260

gacgtgcagt ccgacaccga ggaggccttc atcgacgagg tgcacgaggt gcagcctacc 4320

agcagcggct ccgagatcct ggacgagcag aacgtgatcg agcagcccgg cagctccctg 4380

gccagcaaca ggatcctgac cctgccccag aggaccatca ggggcaagaa caagcactgc 4440

tggtccacct ccaagcccac caggcggagc agggtgtccg ccctgaacat cgtgagaagc 4500

cagaggggcc ccaccaggat gtgcaggaac atctacgacc ccctgctgtg cttcaagctg 4560

ttcttcaccg acgagatcat cagcgagatc gtgaagtgga ccaacgccga gatcagcctg 4620

aagaggcggg agagcatgac ctccgccacc ttcagggaca ccaacgagga cgagatctac 4680

gccttcttcg gcatcctggt gatgaccgcc gtgaggaagg acaaccacat gagcaccgac 4740

gacctgttcg acagatccct gagcatggtg tacgtgagcg tgatgagcag ggacagattc 4800

gacttcctga tcagatgcct gaggatggac gacaagagca tcaggcccac cctgcgggag 4860

aacgacgtgt tcacccccgt gagaaagatc tgggacctgt tcatccacca gtgcatccag 4920

aactacaccc ctggcgccca cctgaccatc gacgagcagc tgctgggctt caggggcagg 4980

tgccccttca gggtctatat ccccaacaag cccagcaagt acggcatcaa gatcctgatg 5040

atgtgcgaca gcggcaccaa gtacatgatc aacggcatgc cctacctggg caggggcacc 5100

cagaccaacg gcgtgccccc gggcgagtac tacgtgaagg agctgtccaa gcccgtccac 5160

ggcagctgca gaaacatcac ctgcgacaac tggttcaccg ccatcccccc ggccaagaac 5220

ctgctgcagg agccctacaa gctgaccatc gtgggcaccg tggccagcaa cgccagagag 5280

atccccgagg tcctgaagaa cagcaggtcc gcccccgtgg gcaccagcat gttctgcttc 5340

gacggccccc tgaccctggt gtcctacaag cccaagcccg ccaagatggt gtacctgctg 5400

tccagctgcg acgaggacgc cagcatcaac gagagcaccg gcaagcccca gatggtgatg 5460

tactacaacc agaccaaggg cggcgtggac accctgaacc agatgtgcag cgtgatgacc 5520

tgcagcagaa agaccaacag ggcccccatg gccctgctgt acggcatgat caacatcgcc 5580

tgcatcaaca gcttcatcat ctacagccac aacgtgagca gcaagggcga gaaggtgcag 5640

agccggaaaa agttcatgcg gaacctgtac atgggcctga cctccagctt catgaggaag 5700

aggctggagg cccccaccct gaagagatac ctgagggaca acatcagcaa catcctgccc 5760

aaagaggtgc ccggcaccag cgacgacagc accgaggagc ccgtgatgaa gaagaggacc 5820

tactgcacct actgtccccc caagatcaga agaaaggcca gcgccgcctg caagaagtgt 5880

aagaaggtca tctgccggga gcacaacatc gacatgtgcc agaggtgtct c 5931

<210> 112
<211> 5931
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered Piggybac R275A/325A/R372A/T560A

<400> 112
atggacaaga agtactccat tgggctcgat atcggcacaa acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggcgcc ctcctgttcg actccgggga aacggccgaa 180

gccacgcggc tcaaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaggtgg atgactcttt cttccatagg 300

ctggaggagt ccttttggt ggaggaggat aaaaagcacg agcgccaccc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420

aagcttgtag acagtactga taaggctgac ttgcggttga tctatctcgc gctggcgcat 480

atgatcaaat ttcggggaca cttcctcatc gagggggacc tgaacccaga caacagcgat 540

gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga agagaacccg 600

atcaacgcat ccggagttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660

cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttggtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780

gatgccaagc ttcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840

cagatcggcg accagtacgc agacctttt ttggcggcaa agaacctgtc agacgccatt 900

ctgctgagtg atattctgcg agtgaacacg gagatcacca aagctccgct gagcgctagt 960

atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc ccttgtcaga 1020

cagcaactgc ctgagaagta caaggaaatt ttcttcgatc agtctaaaaa tggctacgcc 1080

ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagctta acagagaaga tctgttgcgc 1200

aaacagcgca ctttcgacaa tggaagcatc ccccaccaga ttcacctggg cgaactgcac 1260

gctatcctca ggcggcaaga ggatttctac ccctttttga aagataacag ggaaaagatt 1320

gagaaaatcc tcacatttcg gataccctac tatgtaggcc ccctcgcccg gggaaattcc 1380

agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440

gtcgtggata agggggcctc tgcccagtcc ttcatcgaaa ggatgactaa ctttgataaa 1500

aatctgccta acgaaaaggt gcttcctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatacgtc acagaaggga tgagaaagcc agcattcctg 1620

tctggagagc agaagaaagc tatcgtggac ctcctcttca agacgaaccg gaaagttacc 1680

gtgaaacagc tcaaagaaga ctatttcaaa aagattgaat gtttcgactc tgttgaaatc 1740

agcggagtgg aggatcgctt caacgcatcc ctgggaacgt atcacgatct cctgaaaatc 1800

attaaagaca aggacttcct ggacaatgag gagaacgagg acattcttga ggacattgtc 1860

ctcaccctta cgttgtttga agatagggag atgattgaag aacgcttgaa aacttacgct 1920

catctcttcg acgacaaagt catgaaacag ctcaagaggc gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatgggatc cgagacaagc agagtggaaa gacaatcctg 2040

gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100

tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160

cacgagcaca tcgctaatct tgcaggtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280

atcgagatgg cccgagagaa ccaaactacc cagaagggac agaagaacag tagggaaagg 2340

atgaagagga ttgaagaggg tataaaagaa ctggggtccc aaatccttaa ggaacaccca 2400

gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460

gacatgtacg tggatcagga actggacatc aatcggctct ccgactacga cgtggatcat 2520

atcgtgcccc agtcttttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580

gataaaaata gagggaagag tgataacgtc ccctcagaag aagttgtcaa gaaaatgaaa 2640

aattattggc ggcagctgct gaacgccaaa ctgatcacac aacggaagtt cgataatctg 2700

actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760

cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880

aagctggtct cagatttcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940

taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000

tatcccaagc ttgaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060

atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120

aatattatga atttttcaa gaccgagatt acactggcca atggagagat tcggaagcga 3180

ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

gcgacagtcc ggaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300

cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360

gcacgcaaaa aagattggga ccccaagaaa tacggcggat tcgattctcc tacagtcgct 3420

tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480

aaggaactgc tgggcatcac aatcatggag cgatcaagct tcgaaaaaaa ccccatcgac 3540

tttctcgagg cgaaaggata taaagaggtc aaaaaagacc tcatcattaa gcttcccaag 3600

tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

cagaaaggta acgagctggc actgccctct aaatacgtta atttcttgta tctggccagc 3720

cactatgaaa agctcaaagg gtctcccgaa gataatgagc agaagcagct gttcgtggaa 3780

caacacaaac actaccttga tgagatcatc gagcaaataa gcgaattctc caaaagagtg 3840

atcctcgccg acgctaacct cgataaggtg ctttctgctt acaataagca cagggataag 3900

cccatcaggg agcaggcaga aaacattatc cacttgttta ctctgaccaa cttgggcgcg 3960

cctgcagcct tcaagtactt cgacaccacc atagacagaa agcggtacac ctctacaaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcggtgg agattccggt agcgaaacac cggggacttc agaatcggcc 4140

accccggagt ctggcagcag cctggacgac gagcacatcc tgagcgccct gctgcagagc 4200

gacgacgagc tggtcggcga ggacagcgac agcgaggtga gcgaccacgt gagcgaggac 4260

gacgtgcagt ccgacaccga ggaggccttc atcgacgagg tgcacgaggt gcagcctacc 4320

agcagcggct ccgagatcct ggacgagcag aacgtgatcg agcagcccgg cagctccctg 4380

gccagcaaca ggatcctgac cctgccccag aggaccatca ggggcaagaa caagcactgc 4440

tggtccacct ccaagcccac caggcggagc agggtgtccg ccctgaacat cgtgagaagc 4500

cagaggggcc ccaccaggat gtgcaggaac atctacgacc ccctgctgtg cttcaagctg 4560

ttcttcaccg acgagatcat cagcgagatc gtgaagtgga ccaacgccga gatcagcctg 4620

aagaggcggg agagcatgac ctccgccacc ttcagggaca ccaacgagga cgagatctac 4680

gccttcttcg gcatcctggt gatgaccgcc gtgaggaagg acaaccacat gagcaccgac 4740

gacctgttcg acagatccct gagcatggtg tacgtgagcg tgatgagcag ggacagattc 4800

gacttcctga tcagatgcct gaggatggac gacaagagca tcaggcccac cctgcgggag 4860

aacgacgtgt tcacccccgt gagaaagatc tgggacctgt tcatccacca gtgcatccag 4920

aactacaccc ctggcgccca cctgaccatc gacgagcagc tgctgggctt cgccggcagg 4980

tgccccttca gggtctatat ccccaacaag cccagcaagt acggcatcaa gatcctgatg 5040

atgtgcgaca gcggcaccaa gtacatgatc aacggcatgc cctacctggg caggggcacc 5100

cagaccaacg gcgtgccccc ggccgagtac tacgtgaagg agctgtccaa gcccgtccac 5160

ggcagctgca gaaacatcac ctgcgacaac tggttcacca gcatcccccc ggccaagaac 5220

ctgctgcagg agccctacaa gctgaccatc gtgggcaccg tggccagcaa caagagagag 5280

atccccgagg tcctgaagaa cagcaggtcc aggcccgtgg gcaccagcat gttctgcttc 5340

gacggccccc tgaccctggt gtcctacaag cccaagcccg ccaagatggt gtacctgctg 5400

tccagctgcg acgaggacgc cagcatcaac gagagcaccg gcaagcccca gatggtgatg 5460

tactacaacc agaccaaggg cggcgtggac accctggacc agatgtgcag cgtgatgacc 5520

tgcagcagaa agaccaacag gtggcccatg gccctgctgt acggcatgat caacatcgcc 5580

tgcatcaaca gcttcatcat ctacagccac aacgtgagca gcaagggcga gaaggtgcag 5640

agccggaaaa agttcatgcg gaacctgtac atgggcctga cctccagctt catgaggaag 5700

aggctggagg cccccaccct gaagagatac ctgagggaca acatcagcaa catcctgccc 5760

aaagaggtgc ccggcaccag cgacgacagc accgaggagc ccgtgatgaa gaagaggacc 5820

tacgcctact actgtcccag caagatcaga agaaaggcca gcgccagctg caagaagtgt 5880

aagaaggtca tctgccggga gcacaacatc gacatgtgcc agagctgttt c 5931

<210> 113
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA

<400> 113
tatgtacact tctgacccac 20

<210> 114
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA SaCas9

<400> 114
gtatcacaat tccagtgggt 20

<210> 115
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA SaCas9

<400> 115
gtatcacaat tccagtgggt 20

<210> 116
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA SaCas9

<400> 116
ggacaggatc ggcataaccg 20

<210> 117
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA SaCas9

<400> 117
gtgctcgggg ccactaggga 20

<210> 118
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Cpf1 gRNA

<400> 118
acttataatt cactgtatca 20

<210> 119
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Cpf1 gRNA

<400> 119
agcttgatat ccatggaatt 20

<210> 120
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Cpf1 gRNA

<400> 120
tgctcggggc cactagggac 20

<210> 121
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Cpf1 gRNA

<400> 121
cttttgtaaa actttatggt 20

<210> 122
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Cpf1 gRNA

<400> 122
caaaagtaaa tagcccggct 20

<210> 123
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> CjCas9

<400> 123
gccgatcctg tccctagtgg cc 22

<210> 124
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> CjCas9 gRNA

<400> 124
acaattccag tgggtcagaa gt 22

<210> 125
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> CjCas9 gRNA

<400> 125
acacttctga cccactggaa t 21

<210> 126
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> CjCas9 gRNA

<400> 126
gaattccatg gatatcaagc tt 22

<210> 127
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> CjCas9 gRNA

<400> 127
aattccagtg ggtcagaagt gt 22

<210> 128
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX gRNA

<400> 128
tcaagcgcgt gtatgtacac 20

<210> 129
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX gRNA

<400> 129
ggatcggcat aaccggtgaa 20

<210> 130
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX gRNA

<400> 130
tagacatgag gtctatggac 20

<210> 131
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX gRNA

<400> 131
taagcttgat atccatggaa 20

<210> 132
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> CasX gRNA

<400> 132
tataattcac tgtatcacaa 20

<210> 133
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<221> SITE
<222> 1..5
<223> wherein the sequence can be repeated 1 to 50 times

<220>
<223> Linker

<400> 133
Gly Gly Gly Gly Ser

1 5

<210> 134
<211> 5
<212> PRT
<213> Artificial sequence

<220>
<221> SITE
<222> 1..5
<223> wherein the sequence can be repeated 1 to 50 times

<220>
<223> linker

<400> 134
Glu Ala Ala Ala Lys
1 5

<210> 135
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac R275A/N347S/K375A/D450N/S592G

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

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

Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
565 570 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Gly
580 585 590
Cys Phe

<210> 136
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac N347S/D450N/T560A/S573A/F594L

<400> 136
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
245 250 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu

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

<210> 137
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac N347A/D450N

<400> 137
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1   5   10   15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20   25   30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35   40   45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
    50   55   60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65   70   75   80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
                85   90   95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
            100   105   110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
        115   120   125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
    130   135   140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145   150   155   160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
                165   170   175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
            180   185   190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
        195   200   205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
    210   215   220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225   230   235   240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
                245   250   255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
            260   265   270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
        275   280   285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
    290   295   300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305   310   315   320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
                325   330   335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Ser Trp Phe Thr Ser Ile
            340   345   350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val
        355   360   365

```
Gly Thr Val Arg Ser Asn Lys Arg Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asn Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 138
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac R275A/N347S/R372A/D450N/T560A/F594L

<400> 138

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

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

```
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Ala
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Leu
```

<210> 139
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
R202K/R275A/N347S/R372A/D450N/T560A/F594L

<400> 139

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

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

```
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Leu
```

<210> 140
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
R275A/R277A/N347S/R372A/D450N/T560A/S564P/F594L

<400> 140

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

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

<210> 141
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified Hyperactive Piggybac triple mutant
D450N/R372A/K375A/R376A

<400> 141
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
245 250 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
260 265 270
Gly Phe Arg Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
275 280 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
290 295 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305 310 315 320
Gly Val Pro Leu Gly Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
325 330 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile
340 345 350
Pro Leu Ala Lys Asn Leu Leu Gln Glu Pro Tyr Lys Leu Thr Ile Val

```
        355                 360                 365
Gly Thr Val Ala Ser Asn Ala Ala Glu Ile Pro Glu Val Leu Lys Asn
    370                 375                 380
Ser Arg Ser Arg Pro Val Gly Thr Ser Met Phe Cys Phe Asp Gly Pro
385                 390                 395                 400
Leu Thr Leu Val Ser Tyr Lys Pro Lys Pro Ala Lys Met Val Tyr Leu
                405                 410                 415
Leu Ser Ser Cys Asp Glu Asp Ala Ser Ile Asn Glu Ser Thr Gly Lys
            420                 425                 430
Pro Gln Met Val Met Tyr Tyr Asn Gln Thr Lys Gly Gly Val Asp Thr
        435                 440                 445
Leu Asn Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 142
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
R245A/N347S/R372A/D450N/T560A/S564P/S573A/S592G

<400> 142

```
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1               5                   10                  15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
```

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

```
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Ala
545                 550                 555                 560
Tyr Cys Pro Pro Lys Ile Arg Arg Lys Ala Ser Ala Ala Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Gly
            580                 585                 590
Cys Phe
```

<210> 143
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
R277A/G325A/N347A/K375A/D450N/T560A/S564P/S573A/S592G/F594L

<400> 143

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

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

```
545                 550                 555                 560
Tyr Cys Pro Pro Lys Ile Arg Arg Lys Ala Ser Ala Ala Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Gly
            580                 585                 590
Cys Leu
```

<210> 144
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
V34M/R275A/G325A/N347S/S351A/R372A/K375A/D450N/T560A/S564P

<400> 144

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

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

<210> 145
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
G325A/N347S/K375A/D450N/S573A/M589V/S592G

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

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

<210> 146
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac S230N/R277A/N347S/K375A/D450N

<400> 146

```
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1               5                   10                  15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
            20                  25                  30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
        35                  40                  45
```

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

```
Leu Asn Gln Met Cys Ser Val Met Thr Cys Ser Arg Lys Thr Asn Arg
    450                 455                 460
Trp Pro Met Ala Leu Leu Tyr Gly Met Ile Asn Ile Ala Cys Ile Asn
465                 470                 475                 480
Ser Phe Ile Ile Tyr Ser His Asn Val Ser Ser Lys Gly Glu Lys Val
                485                 490                 495
Gln Ser Arg Lys Lys Phe Met Arg Asn Leu Tyr Met Gly Leu Thr Ser
            500                 505                 510
Ser Phe Met Arg Lys Arg Leu Glu Ala Pro Thr Leu Lys Arg Tyr Leu
        515                 520                 525
Arg Asp Asn Ile Ser Asn Ile Leu Pro Lys Glu Val Pro Gly Thr Ser
    530                 535                 540
Asp Asp Ser Thr Glu Glu Pro Val Met Lys Lys Arg Thr Tyr Cys Thr
545                 550                 555                 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
                565                 570                 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
            580                 585                 590
Cys Phe
```

<210> 147
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac T43I/R372A/K375A/A411T/D450N

<400> 147

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

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

545 550 555 560
Tyr Cys Pro Ser Lys Ile Arg Arg Lys Ala Ser Ala Ser Cys Lys Lys
565 570 575
Cys Lys Lys Val Ile Cys Arg Glu His Asn Ile Asp Met Cys Gln Ser
580 585 590
Cys Phe

<210> 148
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
G325A/N347S/S351A/K375A/D450N/S573A/M589V/S592G

<400> 148
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
Tyr Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile

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

<210> 149
<211> 594
<212> PRT
<213> Artificial Sequence

<220>
<223> Modified hyperactive PiggyBac
Y177H/R275A/G325A/K375A/D450N/T560A/S564P/S592G

<400> 149
Met Gly Ser Ser Leu Asp Asp Glu His Ile Leu Ser Ala Leu Leu Gln
1 5 10 15
Ser Asp Asp Glu Leu Val Gly Glu Asp Ser Asp Ser Glu Val Ser Asp
20 25 30
His Val Ser Glu Asp Asp Val Gln Ser Asp Thr Glu Glu Ala Phe Ile
35 40 45
Asp Glu Val His Glu Val Gln Pro Thr Ser Ser Gly Ser Glu Ile Leu
50 55 60
Asp Glu Gln Asn Val Ile Glu Gln Pro Gly Ser Ser Leu Ala Ser Asn
65 70 75 80
Arg Ile Leu Thr Leu Pro Gln Arg Thr Ile Arg Gly Lys Asn Lys His
85 90 95
Cys Trp Ser Thr Ser Lys Pro Thr Arg Arg Ser Arg Val Ser Ala Leu
100 105 110
Asn Ile Val Arg Ser Gln Arg Gly Pro Thr Arg Met Cys Arg Asn Ile
115 120 125
Tyr Asp Pro Leu Leu Cys Phe Lys Leu Phe Phe Thr Asp Glu Ile Ile
130 135 140
Ser Glu Ile Val Lys Trp Thr Asn Ala Glu Ile Ser Leu Lys Arg Arg
145 150 155 160
Glu Ser Met Thr Ser Ala Thr Phe Arg Asp Thr Asn Glu Asp Glu Ile
165 170 175
His Ala Phe Phe Gly Ile Leu Val Met Thr Ala Val Arg Lys Asp Asn
180 185 190
His Met Ser Thr Asp Asp Leu Phe Asp Arg Ser Leu Ser Met Val Tyr
195 200 205
Val Ser Val Met Ser Arg Asp Arg Phe Asp Phe Leu Ile Arg Cys Leu
210 215 220
Arg Met Asp Asp Lys Ser Ile Arg Pro Thr Leu Arg Glu Asn Asp Val
225 230 235 240
Phe Thr Pro Val Arg Lys Ile Trp Asp Leu Phe Ile His Gln Cys Ile
245 250 255
Gln Asn Tyr Thr Pro Gly Ala His Leu Thr Ile Asp Glu Gln Leu Leu
260 265 270
Gly Phe Ala Gly Arg Cys Pro Phe Arg Val Tyr Ile Pro Asn Lys Pro
275 280 285
Ser Lys Tyr Gly Ile Lys Ile Leu Met Met Cys Asp Ser Gly Thr Lys
290 295 300
Tyr Met Ile Asn Gly Met Pro Tyr Leu Gly Arg Gly Thr Gln Thr Asn
305 310 315 320
Gly Val Pro Leu Ala Glu Tyr Tyr Val Lys Glu Leu Ser Lys Pro Val
325 330 335
His Gly Ser Cys Arg Asn Ile Thr Cys Asp Asn Trp Phe Thr Ser Ile

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

<210> 150
<211> 390
<212> DNA
<213> Artificial Sequence

<220>
<223> MCP N55K

<400> 150
atggcttcaa actttactca gttcgtgctc gtggacaatg gtgggacagg ggatgtgaca 60

gtggctcctt ctaatttcgc taatggggtg gcagagtgga tcagctccaa ctcacggagc 120

caggcctaca aggtgacatg cagcgtcagg cagtctagtg cccagaagag aaagtatacc 180

atcaaggtgg aggtccccaa agtggctacc cagacagtgg gcggagtcga actgcctgtc 240

gccgcttgga ggtcctacct gaacatggag ctcactatcc caattttcgc taccaattct 300

gactgtgaac tcatcgtgaa ggcaatgcag gggctcctca aagacggtaa tcctatccct 360

tccgccatcg ccgctaactc aggtatctac 390

<210> 151
<211> 130
<212> PRT
<213> Artificial Sequence

<220>
<223> MCP N55K

<400> 151
Met Ala Ser Asn Phe Thr Gln Phe Val Leu Val Asp Asn Gly Gly Thr
1 5 10 15
Gly Asp Val Thr Val Ala Pro Ser Asn Phe Ala Asn Gly Val Ala Glu
20 25 30
Trp Ile Ser Ser Asn Ser Arg Ser Gln Ala Tyr Lys Val Thr Cys Ser
35 40 45
Val Arg Gln Ser Ser Ala Gln Lys Arg Lys Tyr Thr Ile Lys Val Glu
50 55 60
Val Pro Lys Val Ala Thr Gln Thr Val Gly Gly Val Glu Leu Pro Val
65 70 75 80
Ala Ala Trp Arg Ser Tyr Leu Asn Met Glu Leu Thr Ile Pro Ile Phe
85 90 95
Ala Thr Asn Ser Asp Cys Glu Leu Ile Val Lys Ala Met Gln Gly Leu
100 105 110
Leu Lys Asp Gly Asn Pro Ile Pro Ser Ala Ile Ala Ala Asn Ser Gly
115 120 125
Ile Tyr
130

<210> 152
<211> 162
<212> DNA
<213> Artificial Sequence

<220>
<223> gRNA-MS2(tetraloop)-AAVS1-3 spacer

<400> 152
ggggccacta gggacaggat gttttagagc taggccaaca tgaggatcac ccatgtctgc 60

agggcctagc aagttaaaat aaggctagtc cgttatcaac ttggccaaca tgaggatcac 120

ccatgtctgc agggccaagt ggcaccgagt cggtgctttt tt 162

<210> 153
<211> 162
<212> RNA
<213> Artificial Sequence

<220>
<223> gRNA-MS2(tetraloop)-AAVS1-3 spacer

<400> 153
ggggccacua gggacaggau guuuuagagc uaggccaaca ugaggaucac ccaugucugc 60

agggccuagc aaguuaaaau aaggcuaguc cguuaucaac uuggccaaca ugaggaucac 120

ccaugucugc agggccaagu ggcaccgagu cggugcuuuu uu 162