

DESCRIPTION

Title of Invention: MODIFIED Cas9 PROTEIN, AND USE THEREOF
[Technical Field]
[0001]

5 The present invention relates to a modified Cas9 protein that is miniaturized while maintaining a binding ability to guide RNA, and use thereof.

[Background Art]
[0002]

10 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are known to compose the adaptive immune system that provides acquired resistance against invasive foreign nucleic acids in bacteria and archaea together with Cas (CRISPR-associated) genes. CRISPR frequently originate from phage or
15 plasmid DNA and are composed of 24 bp to 48 bp short, conserved repeat sequences having unique variable DNA sequences referred to as spacers of similar size inserted there between. In addition, a group of genes encoding the Cas protein family is present in the vicinity of the repeat and spacer sequences.

20 [0003]

 In the CRISPR/Cas system, foreign DNA is cleaved into fragments of about 30 bp by the Cas protein family and inserted into CRISPR. Cas1 and Cas2 proteins, which are among the Cas protein family, recognize a base sequence referred to as proto-
25 spacer adjacent motif (PAM) of foreign DNA, cut the upstream, and insert same into the CRISPR sequence of the host, which creates immune memory of bacteria. RNA generated by transcription of a CRISPR sequence including immune memory (referred to as pre-crRNA) is paired with a partially
30 complementary RNA (trans-activating crRNA: tracrRNA) and incorporated into Cas9 protein which is one of the Cas protein family. The pre-crRNA and tracrRNA incorporated into Cas9 are cleaved by RNaseIII to form small RNA fragments (CRISPR-RNAs: crRNAs) containing a foreign sequence (guide sequence), and a
35 Cas9-crRNA-tracrRNA complex is thus formed. The Cas9-crRNA-

tracrRNA complex binds to a foreign invasive DNA complementary to crRNA, and the Cas9 protein, which is an enzyme that cleaves the DNA (nuclease), cleaves the foreign invasive DNA, thereby suppressing and eliminating the function of the DNA that
5 invaded from the outside.

[0004]

In recent years, techniques for applying the CRISPR/Cas system to genome editing have been actively developed. crRNA and tracrRNA are fused, expressed as a tracrRNA-crRNA chimera
10 (hereinafter to be referred to as guide RNA: gRNA), and utilized. Using this, nuclease (RNA-guided nuclease: RGN) is then recruited to cleave genomic DNA at the target site.

[0005]

On the other hand, a system that can regulate the
15 expression level of the target gene can be obtained by fusing a transcriptional regulator such as transcriptional activator (e.g., VP64, VP160 and the like), a transcriptional inhibitor (e.g., KRAB and the like), and the like with a variant (nuclease-null, dCas9) wherein the nuclease of the Cas9 protein
20 in CRISPR/Cas9, which is one of the genome editing systems, is inactivated. For example, to further increase the efficiency of gene activation, it is fused with an activation factor (VP64-Rta) in which two transcriptional activators are linked or an activation factor (VP64-p65-Rta, VPR) in which three
25 transcriptional activators are linked, and the fused dCas9 protein (dCas9-VPR; dCas9 fusion protein) strongly activates expression of the target gene without cleaving DNA.

[0006]

Various variants of Cas9 protein have been created and
30 reported for the purpose of alleviating PAM specificity, modifying (activating/inactivating) nuclease activity, and miniaturizing size (patent documents 1 - 3).

[Document List]

[Patent documents]

35 [0007]

patent document 1: WO 2016/141224A1

patent document 2: WO 2017/010543A1

patent document 3: WO 2018/074979A1

[Summary of Invention]

5 [Technical Problem]

[0008]

Expression of dCas9 fusion protein in vivo requires an expression vector. In gene therapy, adeno-associated virus vector (AAV) is mainly used since it is highly safe and highly
10 efficient. The mountable size of AAV is about 4.4 kb, while the dCas9 protein already occupies about 4 kb, and the constituent of fusion protein is extremely limited for mounting on AAV.

Therefore, the present inventors aim to provide a further
15 miniaturized dCas9 protein variant having DNA binding affinity substantially equivalent to that of a full-length protein.

[Solution to Problem]

[0009]

The present inventors took note of a nuclease-null
20 variant (dSaCas9) of Cas9 derived from *S. aureus* (to be also referred to as SaCas9 in the present description) as a Cas9 protein, and have conducted intensive studies in an attempt to solve the above-mentioned problems. As a result, they have found a specific region that has little influence on the
25 ability to bind to guide RNA even if deleted, and succeeded in producing a miniaturized dSaCas9 protein while maintaining or enhancing the DNA binding affinity by substituting the amino acid at a predetermined position with a specific amino acid, which resulted in the completion of the present invention.

30 Deletion and substitution are also collectively referred to as mutation.

In the present description, dSaCas9 protein before introduction of mutation is sometimes to be referred to as wild-type dSaCas9 (protein), and dSaCas9 protein after
35 introduction of mutation is sometimes to be referred to as

modified dSaCas9 variant (protein).

That is, the present invention provides the following.

[1] A protein having a binding ability to guide RNA and consisting of a sequence comprising an amino acid sequence

5 wherein a continuous deletion region is present between the 721-position and the 755-position in the amino acid sequence shown in SEQ ID NO: 2,

wherein amino acids adjacent to each of the deletion region are linked by a linker consisting of 3 to 10 amino acid residues.

10 [2] The protein of the above-mentioned [1], wherein the linker is a 5 - 9 amino acid length linker composed of glycine (G) and serine (S).

[3] The protein of the above-mentioned [2], wherein the linker is selected from the following:

15 -SGGGS-

-GGSGGS-

-SGSGSGSG-

-SGSGSGSGS-.

[4] The protein of any of the above-mentioned [1] to [3],
20 wherein the deletion region is a region of the 721-position to the 745-position.

[5] The protein of the above-mentioned [4], wherein the protein is shown in SEQ ID NO: 4.

[6] The protein of any of the above-mentioned [1] to [3],
25 wherein the deletion region is a region of the 721-position to the 755-position.

[7] The protein of the above-mentioned [6], wherein the protein is shown in SEQ ID NO: 6.

[8] The protein of any of the above-mentioned [1] to [7],
30 wherein glutamic acid (E) at the 45-position and/or the 163-position are/is substituted with other amino acid(s).

[9] The protein of the above-mentioned [8], wherein said other amino acid is a basic amino acid.

[10] The protein of the above-mentioned [9], wherein the basic
35 amino acid is lysine (K).

[11] The protein of any of the above-mentioned [1] to [10], having identity of 80% or more at a site other than the mutated and/or deleted positions in the SEQ ID NO: 2.

[12] The protein of any of the above-mentioned [1] to [10],
5 wherein one to several amino acids are substituted, deleted, inserted and/or added at a site other than the mutated and/or deleted positions in the SEQ ID NO: 2.

[13] The protein of any of the above-mentioned [1] to [12], wherein a transcriptional regulator protein or domain is linked.

10 [14] The protein of the above-mentioned [13], wherein the transcriptional regulator is a transcriptional activator.

[15] The protein of the above-mentioned [13], wherein the transcriptional regulator is a transcriptional silencer or a transcriptional inhibitor.

15 [16] A nucleic acid encoding the protein of any of the above-mentioned [1] to [15].

[17] A protein-RNA complex provided with the protein of any of the above-mentioned [1] to [16] and a guide RNA comprising a polynucleotide composed of a base sequence complementary to a
20 base sequence located 1 to 20 to 24 bases upstream from a proto-spacer adjacent motif (PAM) sequence in a target double-stranded polynucleotide.

[18] A method for site-specifically modifying a target double-stranded polynucleotide, including

25 a step of mixing and incubating a target double-stranded polynucleotide, a protein and a guide RNA, and

a step of having the aforementioned protein modify the aforementioned target double-stranded polynucleotide at a binding site located upstream of a PAM sequence; wherein,

30 the aforementioned protein is the protein of any of the above-mentioned [1] to [15], and

the aforementioned guide RNA contains a polynucleotide composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from the aforementioned
35 PAM sequence in the aforementioned target double-stranded

polynucleotide.

[19] A method for increasing expression of a target gene in a cell, comprising expressing the protein of the above-mentioned [14] and one or plural guide RNAs for the aforementioned target
5 gene in the aforementioned cell.

[20] A method for decreasing expression of a target gene in a cell, comprising expressing the protein of the above-mentioned [15] and one or plural guide RNAs for the aforementioned target gene in the aforementioned cell.

10 [21] The method of the above-mentioned [19] or [20], wherein the cell is a eukaryotic cell.

[22] The method of the above-mentioned [19] or [20], wherein the cell is a yeast cell, a plant cell or an animal cell.

[Advantageous Effects of Invention]

15 [0010]

According to the present invention, a dSaCas9 protein further miniaturized while having a binding ability to guide RNA can be obtained. The miniaturized dSaCas9 protein makes it possible to mount a larger number of genes into expression
20 vectors limited in capacity.

[Brief Description of Drawings]

[0011]

Fig. 1 is a schematic showing of the structure of wild-type dSaCas9(FL) and dSaCas9 variants (T10 - T11). Both ends
25 of the deletion region (721 - 745 positions) in T10 and both ends of the deletion region (721 - 755 positions) in T11 are linked by a linker peptide (SGGGS).

Fig. 2 is a graph showing the results of suppression of the expression of the MyD88 gene by guide RNA using A2 (SEQ ID
30 NO: 12) as crRNA and a molecule in which wild-type dSaCas9 (FL) or dSaCas9 variant (T10, T11) is linked to the KRAB gene.

Fig. 3 is a graph showing the results of suppression of the expression of the MyD88 gene by guide RNA using A2 (SEQ ID
NO: 12) or A3 (SEQ ID NO: 13) as crRNA and a molecule in which
35 wild-type dSaCas9 (FL) or dSaCas9 variant (T10) is linked to

the KRAB gene.

[Description of Embodiments]

[0012]

The present invention is described below. Unless
5 particularly indicated, the terms used in the present
description have meanings generally used in the pertinent field.

[0013]

<dSaCas9 variant>

The dSaCas9 variant of the present invention is a dSaCas9
10 protein further miniaturized while having a binding ability to
guide RNA. Using the miniaturized dSaCas9 protein, a larger
number of genes can be mounted into a vector.

[0014]

In the present description, "guide RNA" refers to that
15 which mimics the hairpin structure of tracrRNA-crRNA, and
contains in the 5'-terminal region thereof a polynucleotide
composed of a base sequence complementary to a base sequence
located from 1 to preferably 20 to 24 bases, and more
preferably from 1 to preferably 22 to 24 bases, upstream from
20 the PAM sequence in a target double-stranded polynucleotide.
Moreover, guide RNA may contain one or more polynucleotides
composed of a base sequence allowing the obtaining of a hairpin
structure composed of base sequences non-complementary to a
target double-stranded polynucleotide symmetrically arranged so
25 as to form a complementary sequence having a single point as
the axis thereof.

The guide RNA has a function of binding to the dSaCas9
variant of the present invention and leading the protein to a
target DNA. The guide RNA has a sequence at the 5'-terminal
30 which is complementary to the target DNA, and binds to the
target DNA via the complementary sequence, thereby leading the
dSaCas9 variant of the present invention to the target DNA.
Since the dSaCas9 variant does not have a DNA endonuclease, it
does not cleave target DNA though it binds to the target DNA.

35 The guide RNA is designed and prepared based on the

sequence information of the target DNA. Specific examples include sequences such as those used in the Examples.

[0015]

In the present description, the terms "polypeptide",
5 "peptide" and "protein" refer to polymers of amino acid residues and are used interchangeably. In addition, these terms also refer to amino acid polymers in which one or a plurality of amino acid residues are in the form of a chemical analog or modified derivative of the corresponding amino acids
10 present in nature.

[0016]

In the present description, the "basic amino acid" refers to an amino acid having a residue showing basicity in addition to one amino group in a molecule such as lysine, arginine,
15 histidine and the like.

[0017]

In the present description, a "sequence" refers to a nucleotide sequence of an arbitrary length, is a deoxyribonucleotide or ribonucleotide, and may be linear or
20 branched and single-stranded or double-stranded.

In the present description, a "PAM sequence" refers to a sequence present in a target double-stranded polynucleotide that can be recognized by Cas9 protein, and the length and base sequence of the PAM sequence differs according to the bacterial
25 species.

[0018]

Furthermore, in the present description, "N" refers to any one base selected from the group consisting of adenine, cytosine, thymine and guanine, "A" refers to adenine, "G" to
30 guanine, "C" to cytosine, "T" to thymine, "R" to a base having a purine skeleton (adenine or guanine), and "Y" to a base having a pyrimidine skeleton (cytosine or thymine).

[0019]

In the present description, a "polynucleotide" refers to
35 a deoxyribonucleotide or ribonucleotide polymer having linear

or cyclic coordination and may be single-stranded or double-stranded, and should not be interpreted as being restricted with respect to polymer length. In addition, polynucleotides include known analogs of naturally-occurring nucleotides as well as nucleotides in which at least one of the base moieties, sugar moieties and phosphate moieties thereof has been modified (such as a phosphorothioate backbone). In general, an analog of a specific nucleotide has the same base-pairing specificity, and for example, A analogs form base pairs with T.

10 [0020]

The present invention provides a protein having a binding ability to guide RNA and consisting of a sequence comprising an amino acid sequence wherein a continuous deletion region is present between the 721-position and the 755-position in the amino acid sequence shown in SEQ ID NO: 2, wherein amino acids adjacent to each of the deletion region are linked by a linker consisting of 3 to 10 amino acid residues (embodiment 1).

[0021]

20 SEQ ID NO: 2 is a full-length amino acid sequence of dSaCas9 protein. The dSaCas9 protein consists of two lobes of REC lobe (41 - 425 residues) and NUC lobe (1 - 40 residues and 435 - 1053 residues). The two lobes are linked via bridge helix (BH: 41 - 73 residues) rich in arginine and linker loop (426 - 434 residues). The NUC lobe is constituted of RuvC domain (1 - 40, 435 - 480 and 650 - 774 residues), HNH domain (520 - 628 residues), WED domain (788 - 909 residues) and PI domain (910 - 1053 residues). The PI domain is divided into topoisomerase homology (TOPO) domain and C-terminal domain (CTD). The RuvC domain is constituted of 3 separate motifs (RuvC-I - III) and is associated with the HNH domain and PI domain. The HNH domain is linked to RuvC-II and RuvC-III via L1 (481 - 519 residues) linker and L2 (629 - 649 residues) linker, respectively. The WED domain and RucV domain are 30 linked by "phosphate lock" loop (775 - 787 residues).

[0022]

In one embodiment of the present invention, the continuous deletion region present between the 721-position and the 755-position in the amino acid sequence shown in SEQ ID NO: 2 is a region of the 721- to 745-positions (embodiment 1-1).

Examples of the protein of embodiment 1-1 and the gene encoding the same respectively include a protein consisting of the amino acid sequence shown in SEQ ID NO: 4 and a gene consisting of the base sequence shown in SEQ ID NO: 3. The protein corresponds to the dSaCas9 variant T10 described below.

[0023]

In one embodiment of the present invention, the continuous deletion region present between the 721-position and the 755-position in the amino acid sequence shown in SEQ ID NO: 2 is a region of the 721- to 755-positions (embodiment 1-2).

Examples of the protein of embodiment 1-2 and the gene encoding the same respectively include a protein consisting of the amino acid sequence shown in SEQ ID NO: 6 and a gene consisting of the base sequence shown in SEQ ID NO: 5. The protein corresponds to the dSaCas9 variant T11 described below.

[0024]

In another embodiment of the present invention, the present invention provides a protein (embodiment 2) having binding ability to guide RNA and further having mutations at the 45-position and/or the 163-position in addition to the mutations in the aforementioned embodiments 1, 1-1 and 1-2.

The mutation(s) at the 45-position and/or the 163-position are/is specifically substitution of glutamic acid with basic amino acid, preferably with lysine, arginine or histidine, more preferably with lysine.

[0025]

As a method for optionally creating "a continuous deletion region between the 721-position and the 755-position" in the amino acid sequence shown in SEQ ID NO: 2, and a method for "substituting glutamic acid at the 45-position and/or the

163-position with other amino acid" in the amino acid sequence shown in SEQ ID NO: 2, a method including introducing a conventional site-specific mutation into a DNA encoding a predetermined amino acid sequence, and then expressing the DNA
5 by a conventional method can be mentioned. Examples of the method for introducing a site-specific mutation include a method using amber mutation (gapped-duplex method, Nucleic Acids Res., 12, 9441-9456 (1984)), a method by PCR using a primer for mutagenesis, and the like. In addition, it can be
10 easily performed according to the manual and using the Q5 Site-Directed Mutagenesis Kit (NEB).

[0026]

In another embodiment of the present invention, the present invention provides a protein (embodiment 3) that is
15 functionally equivalent to the proteins of the aforementioned embodiments 1, 1-1, 1-2 and 2. To be functionally equivalent to the proteins of the aforementioned embodiments 1, 1-1, 1-2 and 2, the amino acid sequence has identity of 80% or more at a site other than the positions where the mutations have been
20 applied in the SEQ ID NO: 2 in the aforementioned embodiments 1, 1-1, 1-2 and 2 and has a binding ability to guide RNA. When amino acids are increased or decreased due to mutation, the "site other than the position(s) where the mutation(s) has(have) been applied" can be interpreted to mean a "site
25 other than the position(s) corresponding to the position(s) where the mutation(s) has(have) been applied". This identity is preferably 80% or more, more preferably 85% or more, even more preferably 90% or more, particularly preferably 95% or more, and most preferably 99% or more. The amino acid sequence
30 identity can be determined by a method known per se. For example, amino acid sequence identity (%) can be determined using a program conventionally used in the pertinent field (e.g., BLAST, FASTA, etc.) by default. In another aspect, identity (%) is determined by any algorithm known in the
35 pertinent field, such as algorithms of Needleman et al. (1970)

(J. Mol. Biol. 48: 444-453), Myers and Miller (CABIOS, 1988, 4: 11-17) and the like. The algorithm of Needleman et al. is incorporated into the GAP program in the GCG software package (available at www.gcg.com) and the identity (%) can be
5 determined using, for example, any of BLOSUM 62 matrix and PAM250 matrix, as well as gap weight: 16, 14, 12, 10, 8, 6 or 4, and length weight: 1, 2, 3, 4, 5 or 6. The algorithm of Myers and Miller is incorporated into the ALIGN program that is a part of the GCG sequence alignment software package. When the
10 ALIGN program is used to compare amino acid sequences, for example, PAM120 weight residue table, gap length penalty 12, and gap penalty 4 can be used.
[0027]

As a protein functionally equivalent to the proteins of
15 the aforementioned embodiments 1, 1-1, 1-2 and 2, a protein (embodiment 3-1) which comprises one to several amino acids substituted, deleted, inserted and/or added at site(s) other than the positions where the mutations have been applied in the SEQ ID NO: 2 in the aforementioned embodiment 1, 1-1, 1-2 and 2
20 and having the binding ability to guide RNA is provided. When amino acids are increased or decreased due to mutation, the "site other than the position(s) where the mutation(s) has(have) been applied" can be interpreted to mean a "site other than the position(s) corresponding to the position(s)
25 where the mutation(s) has(have) been applied".

As a technique for artificially performing "substitution, deletion, insertion and/or addition of amino acid", for example, a method including applying conventional site specific mutation introduction to DNA encoding a predetermined amino acid
30 sequence, and thereafter expressing the DNA by a conventional method can be mentioned. Examples of the site specific mutation introduction method include a method using amber mutation (gapped duplex method, Nucleic Acids Res., 12, 9441-9456 (1984)), a PCR method using a mutation introduction primer
35 and the like. In addition, it can be easily performed

according to the manual and using the Q5 Site-Directed Mutagenesis Kit (NEB).

The number of the amino acids modified above is at least one residue, specifically one or several, or more than that.

5 Among the aforementioned substitution, deletion, insertion and addition, substitution of amino acid is particularly preferred. The substitution is more preferably substitution with an amino acid having similar properties such as hydrophobicity, charge, pK, and characteristic of steric structure and the like.

10 Examples of the substitution include substitution within the groups of i) glycine, alanine; ii) valine, isoleucine, leucine; iii) aspartic acid, glutamic acid, asparagine, glutamine; iv) serine, threonine; v) lysine, arginine; vi) phenylalanine, tyrosine.

15 [0028]

In the dSaCas9 variant of the present invention, the dSaCas9 variant protein is in a state of being cleaved by the deletion mutation, and the both ends of the deletion region are linked by a linker. That is, in the dSaCas9 variant of the
20 present invention, amino acids each adjacent to the deletion region are linked by a linker consisting of 3 to 10 amino acid residues. Due to the linkage, the dSaCas9 variant of the present invention has a continuous amino acid sequence.

[0029]

25 The linker is not particularly limited as long as it can link both ends of a cleaved protein and does not influence the function thereof. Preferably, it is a group capable of adopting an intrinsically disordered structure that binds to other protein while freely changing its own shape according to
30 the protein, and more preferably a peptide residue having a length of 5 - 9 amino acids which is constituted of glycine (G) and serine (S). Specifically, the following residues can be mentioned.

-SGGGS- (SEQ ID NO: 7)

35 -GGSGGS- (SEQ ID NO: 8)

-SGSGSGSG- (SEQ ID NO: 9)

-SGSGSGSGS- (SEQ ID NO: 10)

In the dSaCas9 variants T10 and T11, -SGGGS- is used as the linker.

5 [0030]

The introduction of linker in each variant can also be performed by a method including performing conventional site-specific mutagenesis on the DNA encoding a predetermined amino acid sequence to insert a base sequence encoding the linker,
10 and thereafter expressing the DNA by a conventional method. Examples of the method for site-specific mutagenesis include methods same as those described above.

[0031]

The dSaCas9 variant in the present embodiment can be
15 produced according to, for example, the method indicated below. First, a host is transformed using a vector containing a nucleic acid that encodes the dSaCas9 variant of the present invention. The n, the host is cultured to express the aforementioned protein. Conditions such as medium composition,
20 culture temperature, duration of culturing or addition of inducing agents can be determined by a person with ordinary skill in the art in accordance with known methods so that the transformant grows and the aforementioned protein is efficiently produced. In addition, in the case of having
25 incorporated a selection marker in the form of an antibiotic resistance gene in an expression vector, the transformant can be selected by adding antibiotic to the medium. The n, dSaCas9 variant of the present invention is obtained by purifying the aforementioned protein expressed by the host according to a
30 method known per se.

There are no particular limitations on the host, and examples thereof include animal cells, plant cells, insect cells and microorganisms such as Escherichia coli, Bacillus subtilis or yeast. Preferred host is an animal cell.

35 [0032]

<dSaCas9 variant-guide RNA complex>

In one embodiment thereof, the present invention provides a protein-RNA complex provided with the protein indicated in the previous section on <dSaCas9 variant> and guide RNA
5 containing a polynucleotide composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from a proto-spacer adjacent motif (PAM) sequence in a target double-stranded polynucleotide.

[0033]

10 The aforementioned protein and the aforementioned guide RNA are able to form a protein-RNA complex by mixing in vitro and in vivo under mild conditions. Mild conditions refer to a temperature and pH of a degree that does not cause proteolysis or denaturation, and the temperature is preferably 4°C to 40°C,
15 while the pH is preferably 4 to 10.

[0034]

In addition, the duration of mixing and incubating the aforementioned protein and the aforementioned guide RNA is preferably 0.5 hr to 1 hr. The complex formed by the
20 aforementioned protein and the aforementioned guide RNA is stable and is able to maintain stability even if allowed to stand for several hours at room temperature.

[0035]

<CRISPR-Cas Vector System>

25 In one embodiment thereof, the present invention provides a CRISPR-Cas vector system provided with a first vector containing a gene encoding a protein indicated in the previous section on <dSaCas9 variant>, and a second vector containing a guide RNA containing a polynucleotide composed of a base
30 sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from PAM sequence in a target double-stranded polynucleotide.

In another embodiment, the present invention provides a CRISPR-Cas vector system in which a gene encoding a protein
35 indicated in the previous section on <dSaCas9 variant>, and a

guide RNA containing a polynucleotide composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from PAM sequence in a target double-stranded polynucleotide are contained in the same vector.

5 [0036]

The guide RNA is suitably designed to contain in the 5'-terminal region thereof a polynucleotide composed of a base sequence complementary to a base sequence located from 1 to 20 to 24 bases, and preferably to 22 to 24 bases, upstream from a
10 PAM sequence in a target double-stranded polynucleotide.

Moreover, the guide RNA may also contain one or more polynucleotides composed of a base sequence allowing the obtaining of a hairpin structure composed of base sequences non-complementary to a target double-stranded polynucleotide
15 symmetrically arranged so as to form a complementary sequence having a single point as the axis thereof.

[0037]

The vector of the present embodiment is preferably an expression vector. Examples of the expression vector that can
20 be used include E. coli-derived plasmids such as pBR322, pBR325, pUC12 or pUC13; B. subtilis-derived plasmids such as pUB110, pTP5 or pC194; yeast-derived plasmids such as pSH19 or pSH15; bacteriophages such as λ phages; viruses such as adenovirus, adeno-associated virus, lentivirus, vaccinia virus, baculovirus
25 or cytomegalovirus; and modified vectors thereof. In view of the activation of gene expression in vivo, a virus vector, particularly an adeno-associated virus, is preferable.

[0038]

In the aforementioned expression vector, there are no
30 particular limitations on the promoters for expression of the aforementioned dSaCas9 variant protein or the aforementioned guide RNA, and examples thereof that can be used include promoters for expression in animal cells such as EF1 α promoter, SR α promoter, SV40 promoter, LTR promoter, cytomegalovirus
35 (CMV) promoter or HSV-tk promoter, promoters for expression in

plant cells such as the 35S promoter of cauliflower mosaic virus (CaMV) or rubber elongation factor (REF) promoter, and promoters for expression in insect cells such as polyhedrin promoter or p10 promoter. The se promoters can be suitably
5 selected according to the aforementioned dSaCas9 variant protein and the aforementioned guide RNA, or the type of cells expressing the aforementioned Cas9 protein and the aforementioned guide RNA.

[0039]

10 The aforementioned expression vector may also further have a multi-cloning site, enhancer, splicing signal, polyadenylation signal, selection marker (drug resistant) and promoter thereof, or replication origin and the like.

[0040]

15 <Method for Site-Specifically Modifying Target Double-Stranded Polynucleotide>

[First Embodiment]

In one embodiment thereof, the present invention provides a method for site-specifically modifying a target double-
20 stranded polynucleotide, provided with:

a step for mixing and incubating a target double-stranded polynucleotide, a protein and a guide RNA, and

a step for having the aforementioned protein modify the aforementioned target double-stranded polynucleotide at a
25 binding site located upstream of a PAM sequence; wherein, the aforementioned target double-stranded polynucleotide has a PAM sequence,

the aforementioned protein is a protein shown in the aforementioned <dSaCas9 variant>, and

30 the aforementioned guide RNA contains a polynucleotide composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from the aforementioned PAM sequence in the aforementioned target double-stranded polynucleotide.

35 [0041]

In the present embodiment, the target double-stranded polynucleotide is not particular limited as long as it has a PAM sequence.

[0042]

5 In the present embodiment, the protein and guide RNA are as described in the aforementioned <dSaCas9 variant>.

[0043]

The following provides a detailed explanation of the method for site-specifically modifying a target double-stranded
10 polynucleotide.

First, the aforementioned protein and the aforementioned guide RNA are mixed and incubated under mild conditions. Mild conditions are as previously described. The incubation time is preferably 0.5 hr to 1 hr. A complex formed by the
15 aforementioned protein and the aforementioned guide RNA is stable and is able to maintain stability even if allowed to stand for several hours at room temperature.

Next, the aforementioned protein and the aforementioned guide RNA form a complex on the aforementioned target double-stranded polynucleotide. The aforementioned protein recognizes
20 PAM sequences, and binds to the aforementioned target double-stranded polynucleotide at a binding site located upstream of the PAM sequence. Successively, a target double-stranded polynucleotide modified to meet the purpose can be obtained in
25 a region determined by the complementary binding of the aforementioned guide RNA and the aforementioned double-stranded polynucleotide.

[0044]

In the present description, the "modification" means that
30 the target double-stranded polynucleotide changes structurally or functionally. For example, structural or functional change of the target double-stranded polynucleotide by the addition of a functional protein or a base sequence can be mentioned. By the modification, the function of the target double-stranded
35 polynucleotide can be altered, deleted, enhanced, or suppressed,

and a new function can be added.

Since the dSaCas9 variant of the present invention does not have endonuclease activity, the protein can bind to the aforementioned target double-stranded polynucleotide at the binding site located upstream of the PAM sequence bond, but stays there and cannot cleave the polynucleotide. Therefore, for example, when a labeled protein such as fluorescent protein (e.g., GFP) and the like is fused to the protein, the labeled protein can be bound to the target double-stranded polynucleotide via dSaCas9 variant protein-guide RNA. By appropriately selecting a substance to be bound to the dSaCas9 variant, various functions can be imparted to the target double-stranded polynucleotide.

Furthermore, transcriptional regulator protein or domain can be linked to the N-terminal or C-terminal of the dSaCas9 variant protein. Examples of the transcriptional regulator or domain thereof include transcriptional activator or domain thereof (e.g., VP64, VP160, NF- κ B p65), transcriptional silencer or domain thereof (e.g., hetero chromatin protein 1(HP1)), and transcriptional inhibitor or domain thereof (e.g., Kruppel-associated box (KRAB), ERF repressor domain (ERD), mSin3A interaction domain (SID)).

It is also possible to link an enzyme that modifies the methylation state of DNA (e.g., DNA methyltransferase (DNMT), TET)), or an enzyme that modifies a histone subunit (e.g., histone acetyltransferase (HAT), histone deacetylase (HDAC), histone methyltransferase, histone demethylase).

[0045]

[Second Embodiment]

In the present embodiment, an expression step may be further provided prior to the incubation step in which the protein described in the previous section on <dSaCas9 variant> and guide RNA are expressed using the previously described CRISPR-Cas vector system.

[0046]

In the expression step of the present embodiment, dSaCas9 variant protein and guide RNA are first expressed using the aforementioned CRISPR-Cas vector system. A specific expression method includes transforming a host using an expression vector
5 containing a gene that encodes dSaCas9 variant protein and an expression vector containing guide RNA, respectively (or expression vector simultaneously containing gene encoding dSaCas9 variant protein and guide RNA). The n, the host is cultured to express the dSaCas9 variant protein and guide RNA.
10 Conditions such as medium composition, culture temperature, duration of culturing or addition of inducing agents can be determined by a person with ordinary skill in the art in accordance with known methods so that the transformant grows and the aforementioned protein is efficiently produced. In
15 addition, in the case of having incorporated a selection marker in the form of an antibiotic resistance gene in the expression vector, the transformant can be selected by adding antibiotic to the medium. The n, the dSaCas9 variant protein and guide RNA are obtained by purifying the dSaCas9 variant protein and
20 guide RNA expressed by the host according to a suitable method.
[0047]

<Method for Site-Specifically Modifying Target Double-Stranded Polynucleotide in Cells>

In one embodiment thereof, the present invention provides
25 a method for site-specifically modifying a target double-stranded polynucleotide in cells, provided with:

a step for introducing the previously described CRISPR-Cas vector system into a cell and expressing protein described in the previous section on <dSaCas9 variant> and guide RNA,

30 a step for having the aforementioned protein bind with the aforementioned target double-stranded polynucleotide at a binding site located upstream of a PAM sequence, and

a step for obtaining a modified target double-stranded polynucleotide in a region determined by complementary binding
35 between the aforementioned guide RNA and the aforementioned

target double-stranded polynucleotide; wherein

the aforementioned guide RNA contains a polynucleotide composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from the aforementioned
5 PAM sequence in the aforementioned target double-stranded polynucleotide.

[0048]

In the expression step of the present embodiment, first, dSaCas9 variant protein and guide RNA are expressed in a cell
10 using the aforementioned CRISPR-Cas vector system.

[0049]

Examples of organisms serving as the origin of the cells targeted for application of the method of the present embodiment include prokaryote, yeast, animal, plant, insect
15 and the like. There are no particular limitations on the aforementioned animals, and examples thereof include, but are not limited to, human, monkey, dog, cat, rabbit, swine, bovine, mouse, rat and the like. In addition, the type of organism serving as the source of the cells can be arbitrarily selected
20 according to the desired type or objective of the target double-stranded polynucleotide.

[0050]

Examples of animal-derived cells targeted for application of the method of the present embodiment include, but are not
25 limited to, germ cells (such as sperm or ova), somatic cells composing the body, stem cells, progenitor cells, cancer cells isolated from the body, cells isolated from the body that are stably maintained outside the body as a result of having become immortalized (cell line), and cells isolated from the body for
30 which the nuclei have been artificially replaced.

[0051]

Examples of somatic cells composing the body include, but are not limited to, cells harvested from arbitrary tissue such as the skin, kidneys, spleen, adrenals, liver, lungs, ovaries,
35 pancreas, uterus, stomach, small intestine, large intestine,

urinary bladder, prostate gland, testes, thymus, muscle, connective tissue, bone, cartilage, vascular tissue, blood, heart, eyes, brain or neural tissue. Specific examples of somatic cells include, but are not limited to, fibroblasts, 5 bone marrow cells, immunocytes (e.g., B lymphocytes, T lymphocytes, neutrophils, macrophages or monocytes etc.), erythrocytes, platelets, osteocytes, bone marrow cells, pericytes, dendritic cells, keratinocytes, adipocytes, mesenchymal cells, epithelial cells, epidermal cells, 10 endothelial cells, intravascular endothelial cells, lymphatic endothelial cells, hepatocytes, pancreatic islet cells (e.g., α cells, β cells, δ cells, ϵ cells or PP cells etc.), chondrocytes, cumulus cells, glia cells, nerve cells (neurons), oligodendrocytes, microglia cells, astrocytes, cardiomyocytes, 15 esophageal cells, muscle cells (e.g., smooth muscle cells or skeletal muscle cells etc.), melanocytes and mononuclear cells, and the like.

[0052]

Stem cells refer to cells having both the ability to 20 self-replicate as well as the ability to differentiate into a plurality of other cell lines. Examples of stem cells include, but are not limited to, embryonic stem cells (ES cells), embryonic tumor cells, embryonic germ stem cells, induced pluripotent stem cells (iPS cells), neural stem cells, 25 hematopoietic stem cells, mesenchymal stem cells, hepatic stem cells, pancreatic stem cells, muscle stem cells, germ stem cells, intestinal stem cells, cancer stem cells and hair follicle stem cells, and the like.

[0053]

30 Cancer cells are cells derived from somatic cells that have acquired reproductive integrity. Examples of the origins of cancer cells include, but are not limited to, breast cancer (e.g., invasive ductal carcinoma, ductal carcinoma in situ, inflammatory breast cancer etc.), prostate cancer (e.g., 35 hormone-dependent prostate cancer or non-hormone dependent

prostate cancer etc.), pancreatic cancer (e.g., pancreatic ductal carcinoma etc.), stomach cancer (e.g., papillary adenocarcinoma, mucinous carcinoma, adenosquamous carcinoma etc.), lung cancer (e.g., non-small cell lung cancer, small
5 cell lung cancer, malignant mesothelioma etc.), colorectal cancer (e.g., gastrointestinal stromal tumor etc.), rectal cancer (e.g., gastrointestinal stromal tumor etc.), colorectal cancer (e.g., familial colorectal cancer, hereditary non-polyposis colorectal cancer, gastrointestinal stromal tumor
10 etc.), small intestinal cancer (e.g., non-Hodgkin's lymphoma, gastrointestinal stromal tumor etc.), esophageal cancer, duodenal cancer, cancer of the tongue, pharyngeal cancer (e.g., nasopharyngeal carcinoma, oropharyngeal carcinoma, hypopharyngeal carcinoma etc.), head and neck cancer, salivary
15 gland cancer, brain tumor (e.g., pineal astrocytoma, pilocytic astrocytoma, diffuse astrocytoma, anaplastic astrocytoma etc.), schwannoma, liver cancer (e.g., primary liver cancer, extrahepatic bile duct cancer etc.), kidney cancer (e.g., renal cell carcinoma, transitional cell carcinoma of the renal pelvis
20 and ureter etc.), gall bladder cancer, bile duct cancer, pancreatic cancer, endometrial carcinoma, cervical cancer, ovarian cancer (e.g., epithelial ovarian cancer, extragonadal germ cell tumor, ovarian germ cell tumor, ovarian low malignant potential tumor etc.), bladder cancer, urethral cancer, skin
25 cancer (e.g., intraocular (ocular) melanoma, Merkel cell carcinoma etc.), Hemangioma, malignant lymphoma (e.g., reticulum cell sarcoma, lymphosarcoma, Hodgkin's etc.), melanoma (malignant melanoma), thyroid cancer (e.g., medullary thyroid cancer etc.), parathyroid cancer, nasal cancer,
30 paranasal cancer, bone tumor (e.g., osteosarcoma, Ewing's tumor, uterine sarcoma, soft tissue sarcoma etc.), metastatic medulloblastoma, vascular fibroma, protuberant dermatofibrosarcoma, retinal sarcoma, penile cancer, testicular cancer, pediatric solid tumor (e.g., Wilms tumor or pediatric
35 kidney tumor etc.), Kaposi's sarcoma, AIDS-induced Kaposi's

sarcoma, maxillary sinus tumor, fibrous histiocytoma, leiomyosarcoma, rhabdomyosarcoma, chronic myeloproliferative disease and leukemia (e.g., acute myeloid leukemia, acute lymphoblastic leukemia etc.).

5 [0054]

Cell lines refer to cells that have acquired reproductive integrity through artificial manipulation ex vivo. Examples of cell lines include, but are not limited to, HCT116, Huh7, HEK293 (human embryonic kidney cells), HeLa (human cervical
10 cancer cell line), HepG2 (human liver cancer cell line), UT7/TPO (human leukemia cell line), CHO (Chinese hamster ovary cell line), MDCK, MDBK, BHK, C-33A, HT-29, AE-1, 3D9, NsO/1, Jurkat, NIH3T3, PC12, S2, Sf9, Sf21, High Five and Vero.

[0055]

15 Introduction of the CRISPR-Cas vector system into cells can be carried out using a method suitable for the viable cells used, and examples thereof include electroporation method, heat shock method, calcium phosphate method, lipofection method, DEAE dextran method, microinjection method, particle gun method,
20 methods using viruses, and methods using commercially available transfection reagents such as FuGENE (registered trade mark) 6 Transfection Reagent (manufactured by Roche), Lipofectamine 2000 Reagent (manufactured by Invitrogen Corp.), Lipofectamine LTX Reagent (manufactured by Invitrogen Corp.) or Lipofectamine
25 3000 Reagent (manufactured by Invitrogen Corp.).

[0056]

The subsequent modification step is the same as the method shown in the aforementioned <Method for Site-Specifically Modifying Target Double-Stranded nucleotide>
30 [First Embodiment].

By the modification of the target double-stranded polynucleotide in this embodiment, cells with modified target double-stranded polynucleotide can be obtained.

[0057]

35 While the present invention is explained in more detail

in the following by referring to Examples, they do not limit the scope of the present invention.

[Example]

[0058]

5 Example 1: Evaluation of DNA binding affinity of dSaCas9 variant
(Method)

1. Cloning

Using NEB Q5 Site-Directed Mutagenesis Kit, a
10 predetermined deletion region was made in dSaCas9 gene, a gene encoding a linker was introduced, and a KRAB gene as a transcriptional regulator was fused to produce various dSaCas9s (Fig. 1). The expression suppression activity of these dSaCas9s was examined using MYD88 gene. All gene constructs of
15 dSaCas9 were incorporated into pX601 vectors (F. Ann Ran et al., Nature 2015; 520(7546); pp.186-191). Each vector expresses the following dSaCas9 constructs (FL, T10, T11) and sgRNA.

[dSaCas9 variants]

FL; wild-type dSaCas9 (SEQ ID NO: 1, SEQ ID NO: 2)

20 T10; deletion variant (SEQ ID NO: 3, SEQ ID NO: 4) having a deletion region (721- to 745-positions) in wild-type dSaCas9
T11; deletion variant (SEQ ID NO: 5, SEQ ID NO: 6) having a deletion region (721- to 755-positions) in wild-type dSaCas9

Both ends of the respective deletion regions in T10 and
25 T11 are linked by a linker peptide.

dSaCas9 construct is expressed as a fusion protein of dSaCas9-KRAB-P2A-sfGFP or dSaCas9-VR-P2A-sfGFP.

[crRNA sequences]

C1; ACGGAGGCUAAGCGUCGCAA (SEQ ID NO: 11)

30 A2; GGAGCCACAGUUCUCCACGG (SEQ ID NO: 12)

A3; CUCUACCCUUGAGGUCUCGAG (SEQ ID NO: 13)

The above-mentioned crRNA sequence was fused with the following tracrRNA sequence to form sgRNA to be expressed from the vector.

35 [tracrRNA sequence]

GUUUUAGUACUCUGGAAACAGAAUCUACUAAAACAAGGC AAAAUGCCGUGUUUAUCACGUCAA
CUUGUUGGCGAGAUUUUUUUU (SEQ ID NO: 14)

2. Cell transfection

HEK293FT cells were seeded in a 24-well plate at a
5 density of 75,000 cells per well 24 hr before transfection and
cultured in DMEM medium supplemented with 10% FBS, 2 mM fresh
L-glutamine, 1 mM sodium pyruvate and non-essential amino acid.
Each plasmid (500 ng) was transfected into cells according to
the manual and using 1.5 µl of Lipofectamine 2000 (Life
10 technologies). For the gene expression analysis, the cells
were recovered at 48-72 hr after transfection, dissolved in RLT
buffer (Qiagen), and the total RNA was extracted using RNeasy
kit (Qiagen).

3. Gene expression analysis

15 For the Taqman analysis, cDNA was prepared from 1.5 µg of
the total RNA by using 20 µl volume TaqMan™ High-Capacity RNA-
to-cDNA Kit (Applied Biosystems). The prepared cDNA was
diluted 20-fold and 6.33 µl was used for each Taqman reaction.
The Taqman primers and probes for the MYD88 gene and HPRT gene
20 were obtained from Applied Biosystems. In Roche LightCycler 96
or LightCycler 480, the Taqman reaction was run using Taqman
gene expression master mix (ThermoFisher), and the analysis was
performed using LightCycler 96 analysis software.

Taqman probe product IDs:

25 MYD88: Hs01573837_g1 (FAM)
HPRT: Hs99999909_m1 (FAM, VIC)

Taqman QPCR condition:

Step 1; 95°C 10 min
30 Step 2; 95°C 15 sec
Step 3; 60°C 30 sec
Repeat Step 2 and 3; 40 times
[0059]
(Results)

35 HEK293FT cells were transfected with a plasmid vector

expressing pX601-dSaCas9 (FL, T10 or T11)-KRAB-P2A-sfGFP and sgRNA-A2. The cells transfected with a vector expressing sgRNA-C1 instead of sgRNA-A2 were used as a negative control. The transfected cells were harvested 3 days later and RNA was isolated. The expression of the MYD88 gene was analyzed by Tagman assay and standardized at the expression level of the HPRT gene. The results of the relative expression level of the MYD88 gene are shown in Fig. 2.

When compared with the control, the MYD88 gene expression level in the dSaCas9 variant of the present invention was as low as that of the wild-type dSaCas9. From the results, it was shown that the binding ability to the guide RNA, and further, the DNA binding affinity, were maintained even though the dSaCas9 variant of the present invention has a deletion region and is smaller than the full-length dSaCas9.

The binding affinity of T10 which showed particularly high DNA binding affinity in the above-mentioned results was confirmed using a different guide RNA. HEK293FT cells were transfected with a plasmid vector expressing pX601-dSaCas9 (FL or T10)-KRAB-P2A-sfGFP and sgRNA-A2 or sgRNA-A3. The cells transfected with a vector not expressing dSaCas9 variant were used as a negative control. The transfected cells were harvested 3 days later and RNA was isolated. The expression of the MYD88 gene was analyzed by Tagman assay and standardized at the expression level of the HPRT gene. The results of the relative expression level of the MYD88 gene are shown in Fig. 3. Even when sgRNA-A3 was used, an effect similar to that of sgRNA-A2 was confirmed.

When compared with the control, the MYD88 gene expression level in the dSaCas9 variant of the present invention was as low as that of the wild-type dSaCas9. From the results, it was shown that the binding ability to the guide RNA, and further, the DNA binding affinity, were maintained even though the dSaCas9 variant of the present invention has a deletion region and is smaller than the full-length dSaCas9.

[Industrial Applicability]

[0060]

According to the present invention, a dSaCas9 protein
that is miniaturized while maintaining a DNA binding ability
5 can be obtained. Use of the miniaturized dSaCas9 protein makes
it possible to mount many genes into vectors, and thus provides
various genome editing techniques.

[0061]

This application is based on a provisional patent
10 application No. 62/749,855 filed in the US (filing date:
October 24, 2018), the contents of which are incorporated in
full herein.

SEQUENCE LISTING

<110> Modalis Therapeutics Corporation
 <120> Modified Cas9 protein and use thereof
 <130> 092961
 <150> US62/749855
 <151> 2018-10-24
 <160> 14
 <170> PatentIn version 3.5
 <210> 1
 <211> 3162
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> CDS
 <222> (1).. (3162)

<400> 1
 atg aag cgg aac tac atc ctg ggc ctg gcc atc ggc atc acc agc gtg 48
 Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
 1 5 10 15
 ggc tac ggc atc atc gac tac gag aca cgg gac gtg atc gat gcc ggc 96
 Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
 20 25 30
 gtg cgg ctg ttc aaa gag gcc aac gtg gaa aac aac gag ggc agg cgg 144
 Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
 35 40 45
 agc aag aga ggc gcc aga agg ctg aag cgg cgg agg cgg cat aga atc 192
 Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
 50 55 60
 cag aga gtg aag aag ctg ctg ttc gac tac aac ctg ctg acc gac cac 240
 Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
 65 70 75 80
 agc gag ctg agc ggc atc aac ccc tac gag gcc aga gtg aag ggc ctg 288
 Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
 85 90 95
 agc cag aag ctg agc gag gaa gag ttc tct gcc gcc ctg ctg cac ctg 336
 Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
 100 105 110
 gcc aag aga aga ggc gtg cac aac gtg aac gag gtg gaa gag gac acc 384
 Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
 115 120 125

ggc aac gag ctg tcc acc aaa gag cag atc agc cgg aac agc aag gcc Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala 130 135 140	432
ctg gaa gag aaa tac gtg gcc gaa ctg cag ctg gaa cgg ctg aag aaa Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys 145 150 155 160	480
gac ggc gaa gtg cgg ggc agc atc aac aga ttc aag acc agc gac tac Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr 165 170 175	528
gtg aaa gaa gcc aaa cag ctg ctg aag gtg cag aag gcc tac cac cag Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln 180 185 190	576
ctg gac cag agc ttc atc gac acc tac atc gac ctg ctg gaa acc cgg Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg 195 200 205	624
cgg acc tac tat gag gga cct ggc gag ggc agc ccc ttc ggc tgg aag Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys 210 215 220	672
gac atc aaa gaa tgg tac gag atg ctg atg ggc cac tgc acc tac ttc Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe 225 230 235 240	720
ccc gag gaa ctg cgg agc gtg aag tac gcc tac aac gcc gac ctg tac Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr 245 250 255	768
aac gcc ctg aac gac ctg aac aat ctc gtg atc acc agg gac gag aac Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn 260 265 270	816
gag aag ctg gaa tat tac gag aag ttc cag atc atc gag aac gtg ttc Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe 275 280 285	864
aag cag aag aag aag ccc acc ctg aag cag atc gcc aaa gaa atc ctc Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu 290 295 300	912
gtg aac gaa gag gat att aag ggc tac aga gtg acc agc acc ggc aag Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys 305 310 315 320	960
ccc gag ttc acc aac ctg aag gtg tac cac gac atc aag gac att acc Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr 325 330 335	1008
gcc cgg aaa gag att att gag aac gcc gag ctg ctg gat cag att gcc Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala 340 345 350	1056
aag atc ctg acc atc tac cag agc agc gag gac atc cag gaa gaa ctg Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu	1104

355					360					365						
acc	aat	ctg	aac	tcc	gag	ctg	acc	cag	gaa	gag	atc	gag	cag	atc	tct	1152
Thr	Asn	Leu	Asn	Ser	Glu	Leu	Thr	Gln	Glu	Glu	Ile	Glu	Gln	Ile	Ser	
	370					375					380					
aat	ctg	aag	ggc	tat	acc	ggc	acc	cac	aac	ctg	agc	ctg	aag	gcc	atc	1200
Asn	Leu	Lys	Gly	Tyr	Thr	Gly	Thr	His	Asn	Leu	Ser	Leu	Lys	Ala	Ile	
385					390					395					400	
aac	ctg	atc	ctg	gac	gag	ctg	tgg	cac	acc	aac	gac	aac	cag	atc	gct	1248
Asn	Leu	Ile	Leu	Asp	Glu	Leu	Trp	His	Thr	Asn	Asp	Asn	Gln	Ile	Ala	
				405					410					415		
atc	ttc	aac	cgg	ctg	aag	ctg	gtg	ccc	aag	aag	gtg	gac	ctg	tcc	cag	1296
Ile	Phe	Asn	Arg	Leu	Lys	Leu	Val	Pro	Lys	Lys	Val	Asp	Leu	Ser	Gln	
			420					425					430			
cag	aaa	gag	atc	ccc	acc	acc	ctg	gtg	gac	gac	ttc	atc	ctg	agc	ccc	1344
Gln	Lys	Glu	Ile	Pro	Thr	Thr	Leu	Val	Asp	Asp	Phe	Ile	Leu	Ser	Pro	
		435					440					445				
gtc	gtg	aag	aga	agc	ttc	atc	cag	agc	atc	aaa	gtg	atc	aac	gcc	atc	1392
Val	Val	Lys	Arg	Ser	Phe	Ile	Gln	Ser	Ile	Lys	Val	Ile	Asn	Ala	Ile	
	450					455					460					
atc	aag	aag	tac	ggc	ctg	ccc	aac	gac	atc	att	atc	gag	ctg	gcc	cgc	1440
Ile	Lys	Lys	Tyr	Gly	Leu	Pro	Asn	Asp	Ile	Ile	Ile	Glu	Leu	Ala	Arg	
465					470					475					480	
gag	aag	aac	tcc	aag	gac	gcc	cag	aaa	atg	atc	aac	gag	atg	cag	aag	1488
Glu	Lys	Asn	Ser	Lys	Asp	Ala	Gln	Lys	Met	Ile	Asn	Glu	Met	Gln	Lys	
				485					490					495		
cgg	aac	cgg	cag	acc	aac	gag	cgg	atc	gag	gaa	atc	atc	cgg	acc	acc	1536
Arg	Asn	Arg	Gln	Thr	Asn	Glu	Arg	Ile	Glu	Glu	Ile	Ile	Arg	Thr	Thr	
			500					505					510			
ggc	aaa	gag	aac	gcc	aag	tac	ctg	atc	gag	aag	atc	aag	ctg	cac	gac	1584
Gly	Lys	Glu	Asn	Ala	Lys	Tyr	Leu	Ile	Glu	Lys	Ile	Lys	Leu	His	Asp	
		515					520					525				
atg	cag	gaa	ggc	aag	tgc	ctg	tac	agc	ctg	gaa	gcc	atc	cct	ctg	gaa	1632
Met	Gln	Glu	Gly	Lys	Cys	Leu	Tyr	Ser	Leu	Glu	Ala	Ile	Pro	Leu	Glu	
	530					535					540					
gat	ctg	ctg	aac	aac	ccc	ttc	aac	tat	gag	gtg	gac	cac	atc	atc	ccc	1680
Asp	Leu	Leu	Asn	Asn	Pro	Phe	Asn	Tyr	Glu	Val	Asp	His	Ile	Ile	Pro	
545					550					555					560	
aga	agc	gtg	tcc	ttc	gac	aac	agc	ttc	aac	aac	aag	gtg	ctc	gtg	aag	1728
Arg	Ser	Val	Ser	Phe	Asp	Asn	Ser	Phe	Asn	Asn	Lys	Val	Leu	Val	Lys	
				565					570					575		
cag	gaa	gaa	gcc	agc	aag	aag	ggc	aac	cgg	acc	cca	ttc	cag	tac	ctg	1776
Gln	Glu	Glu	Ala	Ser	Lys	Lys	Gly	Asn	Arg	Thr	Pro	Phe	Gln	Tyr	Leu	
			580					585					590			

agc Ser	agc Ser	agc Ser 595	gac Asp	agc Ser	aag Lys	atc Ile	agc Ser 600	tac Tyr	gaa Glu	acc Thr	ttc Phe 605	aag Lys	aag Lys	cac His	atc Ile	1824
ctg Leu 610	aat Asn	ctg Leu	gcc Ala	aag Lys	ggc Gly	aag Lys 615	ggc Gly	aga Arg	atc Ile	agc Ser	aag Lys 620	acc Thr	aag Lys	aaa Lys	gag Glu	1872
tat Tyr 625	ctg Leu	ctg Leu	gaa Glu	gaa Glu	cgg Arg 630	gac Asp	atc Ile	aac Asn	agg Arg	ttc Phe 635	tcc Ser	gtg Val	cag Gln	aaa Lys	gac Asp 640	1920
ttc Phe	atc Ile	aac Asn	cgg Arg	aac Asn 645	ctg Leu	gtg Val	gat Asp	acc Thr	aga Arg 650	tac Tyr	gcc Ala	acc Thr	aga Arg	ggc Gly 655	ctg Leu	1968
atg Met	aac Asn	ctg Leu 660	ctg Leu	cgg Arg	agc Ser	tac Tyr	ttc Phe	aga Arg 665	gtg Val	aac Asn	aac Asn	ctg Leu	gac Asp 670	gtg Val	aaa Lys	2016
gtg Val	aag Lys	tcc Ser 675	atc Ile	aat Asn	ggc Gly	ggc Gly	ttc Phe 680	acc Thr	agc Ser	ttt Phe	ctg Leu	cgg Arg 685	cgg Arg	aag Lys	tgg Trp	2064
aag Lys 690	ttt Phe	aag Lys	aaa Lys	gag Glu	cgg Arg	aac Asn 695	aag Lys	ggg Gly	tac Tyr	aag Lys	cac His 700	cac His	gcc Ala	gag Glu	gac Asp	2112
gcc Ala 705	ctg Leu	atc Ile	att Ile	gcc Ala	aac Asn 710	gcc Ala	gat Asp	ttc Phe	atc Ile	ttc Phe 715	aaa Lys	gag Glu	tgg Trp	aag Lys	aaa Lys 720	2160
ctg Leu	gac Asp	aag Lys	gcc Ala	aaa Lys 725	aaa Lys	gtg Val	atg Met	gaa Glu	aac Asn 730	cag Gln	atg Met	ttc Phe	gag Glu	gaa Glu 735	aag Lys	2208
cag Gln	gcc Ala	gag Glu	agc Ser 740	atg Met	ccc Pro	gag Glu	atc Ile	gaa Glu 745	acc Thr	gag Glu	cag Gln	gag Glu	tac Tyr 750	aaa Lys	gag Glu	2256
atc Ile	ttc Phe	atc Ile 755	acc Thr	ccc Pro	cac His	cag Gln	atc Ile 760	aag Lys	cac His	att Ile	aag Lys	gac Asp 765	ttc Phe	aag Lys	gac Asp	2304
tac Tyr 770	aag Lys	tac Tyr	agc Ser	cac His	cgg Arg	gtg Val 775	gac Asp	aag Lys	aag Lys	cct Pro	aat Asn 780	aga Arg	gag Glu	ctg Leu	att Ile	2352
aac Asn 785	gac Asp	acc Thr	ctg Leu	tac Tyr	tcc Ser 790	acc Thr	cgg Arg	aag Lys	gac Asp	gac Asp 795	aag Lys	ggc Gly	aac Asn	acc Thr	ctg Leu 800	2400
atc Ile	gtg Val	aac Asn	aat Asn	ctg Leu 805	aac Asn	ggc Gly	ctg Leu	tac Tyr	gac Asp 810	aag Lys	gac Asp	aat Asn	gac Asp	aag Lys 815	ctg Leu	2448
aaa Lys	aag Lys	ctg Leu	atc Ile	aac Asn	aag Lys	agc Ser	ccc Pro	gaa Glu	aag Lys	ctg Leu	ctg Leu	atg Met	tac Tyr	cac His	cac His	2496

820	825	830	
gac ccc cag acc tac cag aaa ctg aag ctg att atg gaa cag tac ggc Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly 835 840 845			2544
gac gag aag aat ccc ctg tac aag tac tac gag gaa acc ggg aac tac Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr 850 855 860			2592
ctg acc aag tac tcc aaa aag gac aac ggc ccc gtg atc aag aag att Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile 865 870 875 880			2640
aag tat tac ggc aac aaa ctg aac gcc cat ctg gac atc acc gac gac Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp 885 890 895			2688
tac ccc aac agc aga aac aag gtc gtg aag ctg tcc ctg aag ccc tac Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr 900 905 910			2736
aga ttc gac gtg tac ctg gac aat ggc gtg tac aag ttc gtg acc gtg Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val 915 920 925			2784
aag aat ctg gat gtg atc aaa aaa gaa aac tac tac gaa gtg aat agc Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser 930 935 940			2832
aag tgc tat gag gaa gct aag aag ctg aag aag atc agc aac cag gcc Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala 945 950 955 960			2880
gag ttt atc gcc tcc ttc tac aac aac gat ctg atc aag atc aac ggc Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly 965 970 975			2928
gag ctg tat aga gtg atc ggc gtg aac aac gac ctg ctg aac cgg atc Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile 980 985 990			2976
gaa gtg aac atg atc gac atc acc tac cgc gag tac ctg gaa aac atg Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met 995 1000 1005			3024
aac gac aag agg ccc ccc agg atc att aag aca atc gcc tcc aag Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys 1010 1015 1020			3069
acc cag agc att aag aag tac agc aca gac att ctg ggc aac ctg Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu 1025 1030 1035			3114
tat gaa gtg aaa tct aag aag cac cct cag atc atc aaa aag ggc Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly 1040 1045 1050			3159

<210> 2
 <211> 1053
 <212> PRT
 <213> Staphylococcus aureus

<400> 2

Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
 1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
 20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
 35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
 50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
 65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
 85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
 100 105 110

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

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
 130 135 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
 145 150 155 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
 165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
225 230 235 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
340 345 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
355 360 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
450 455 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
545 550 555 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575

Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
690 695 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705 710 715 720

Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
725 730 735

Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
740 745 750

Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
755 760 765

Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
770 775 780

Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
785 790 795 800

Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
805 810 815

Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
820 825 830

Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
835 840 845

Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
850 855 860

Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
865 870 875 880

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
885 890 895

Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
 900 905 910

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
 915 920 925

Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
 930 935 940

Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
 945 950 955 960

Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
 965 970 975

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
 980 985 990

Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
 995 1000 1005

Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
 1010 1015 1020

Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
 1025 1030 1035

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1040 1045 1050

<210> 3
 <211> 3102
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> CDS
 <222> (1).. (3102)

<400> 3
 atg aag cgg aac tac atc ctg ggc ctg gcc atc ggc atc acc agc gtg 48
 Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
 1 5 10 15

ggc tac ggc atc atc gac tac gag aca cgg gac gtg atc gat gcc ggc 96

Gly	Tyr	Gly	Ile	Ile	Asp	Tyr	Glu	Thr	Arg	Asp	Val	Ile	Asp	Ala	Gly	
			20					25					30			
gtg	cgg	ctg	ttc	aaa	gag	gcc	aac	gtg	gaa	aac	aac	gag	ggc	agg	cgg	144
Val	Arg	Leu	Phe	Lys	Glu	Ala	Asn	Val	Glu	Asn	Asn	Glu	Gly	Arg	Arg	
		35					40					45				
agc	aag	aga	ggc	gcc	aga	agg	ctg	aag	cgg	cgg	agg	cgg	cat	aga	atc	192
Ser	Lys	Arg	Gly	Ala	Arg	Arg	Leu	Lys	Arg	Arg	Arg	Arg	His	Arg	Ile	
	50					55					60					
cag	aga	gtg	aag	aag	ctg	ctg	ttc	gac	tac	aac	ctg	ctg	acc	gac	cac	240
Gln	Arg	Val	Lys	Lys	Leu	Leu	Phe	Asp	Tyr	Asn	Leu	Leu	Thr	Asp	His	
65					70					75					80	
agc	gag	ctg	agc	ggc	atc	aac	ccc	tac	gag	gcc	aga	gtg	aag	ggc	ctg	288
Ser	Glu	Leu	Ser	Gly	Ile	Asn	Pro	Tyr	Glu	Ala	Arg	Val	Lys	Gly	Leu	
				85					90					95		
agc	cag	aag	ctg	agc	gag	gaa	gag	ttc	tct	gcc	gcc	ctg	ctg	cac	ctg	336
Ser	Gln	Lys	Leu	Ser	Glu	Glu	Glu	Phe	Ser	Ala	Ala	Leu	Leu	His	Leu	
			100					105					110			
gcc	aag	aga	aga	ggc	gtg	cac	aac	gtg	aac	gag	gtg	gaa	gag	gac	acc	384
Ala	Lys	Arg	Arg	Gly	Val	His	Asn	Val	Asn	Glu	Val	Glu	Glu	Asp	Thr	
		115					120					125				
ggc	aac	gag	ctg	tcc	acc	aaa	gag	cag	atc	agc	cgg	aac	agc	aag	gcc	432
Gly	Asn	Glu	Leu	Ser	Thr	Lys	Glu	Gln	Ile	Ser	Arg	Asn	Ser	Lys	Ala	
	130					135					140					
ctg	gaa	gag	aaa	tac	gtg	gcc	gaa	ctg	cag	ctg	gaa	cgg	ctg	aag	aaa	480
Leu	Glu	Glu	Lys	Tyr	Val	Ala	Glu	Leu	Gln	Leu	Glu	Arg	Leu	Lys	Lys	
145					150					155					160	
gac	ggc	gaa	gtg	cgg	ggc	agc	atc	aac	aga	ttc	aag	acc	agc	gac	tac	528
Asp	Gly	Glu	Val	Arg	Gly	Ser	Ile	Asn	Arg	Phe	Lys	Thr	Ser	Asp	Tyr	
				165					170					175		
gtg	aaa	gaa	gcc	aaa	cag	ctg	ctg	aag	gtg	cag	aag	gcc	tac	cac	cag	576
Val	Lys	Glu	Ala	Lys	Gln	Leu	Leu	Lys	Val	Gln	Lys	Ala	Tyr	His	Gln	
			180					185					190			
ctg	gac	cag	agc	ttc	atc	gac	acc	tac	atc	gac	ctg	ctg	gaa	acc	cgg	624
Leu	Asp	Gln	Ser	Phe	Ile	Asp	Thr	Tyr	Ile	Asp	Leu	Leu	Glu	Thr	Arg	
		195					200					205				
cgg	acc	tac	tat	gag	gga	cct	ggc	gag	ggc	agc	ccc	ttc	ggc	tgg	aag	672
Arg	Thr	Tyr	Tyr	Glu	Gly	Pro	Gly	Glu	Gly	Ser	Pro	Phe	Gly	Trp	Lys	
	210					215					220					
gac	atc	aaa	gaa	tgg	tac	gag	atg	ctg	atg	ggc	cac	tgc	acc	tac	ttc	720
Asp	Ile	Lys	Glu	Trp	Tyr	Glu	Met	Leu	Met	Gly	His	Cys	Thr	Tyr	Phe	
225					230					235					240	
ccc	gag	gaa	ctg	cgg	agc	gtg	aag	tac	gcc	tac	aac	gcc	gac	ctg	tac	768
Pro	Glu	Glu	Leu	Arg	Ser	Val	Lys	Tyr	Ala	Tyr	Asn	Ala	Asp	Leu	Tyr	
				245					250					255		

aac gcc ctg aac gac ctg aac aat ctc gtg atc acc agg gac gag aac Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn 260 265 270	816
gag aag ctg gaa tat tac gag aag ttc cag atc atc gag aac gtg ttc Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe 275 280 285	864
aag cag aag aag aag ccc acc ctg aag cag atc gcc aaa gaa atc ctc Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu 290 295 300	912
gtg aac gaa gag gat att aag ggc tac aga gtg acc agc acc ggc aag Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys 305 310 315 320	960
ccc gag ttc acc aac ctg aag gtg tac cac gac atc aag gac att acc Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr 325 330 335	1008
gcc cgg aaa gag att att gag aac gcc gag ctg ctg gat cag att gcc Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala 340 345 350	1056
aag atc ctg acc atc tac cag agc agc gag gac atc cag gaa gaa ctg Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu 355 360 365	1104
acc aat ctg aac tcc gag ctg acc cag gaa gag atc gag cag atc tct Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser 370 375 380	1152
aat ctg aag ggc tat acc ggc acc cac aac ctg agc ctg aag gcc atc Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile 385 390 395 400	1200
aac ctg atc ctg gac gag ctg tgg cac acc aac gac aac cag atc gct Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala 405 410 415	1248
atc ttc aac cgg ctg aag ctg gtg ccc aag aag gtg gac ctg tcc cag Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln 420 425 430	1296
cag aaa gag atc ccc acc acc ctg gtg gac gac ttc atc ctg agc ccc Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro 435 440 445	1344
gtc gtg aag aga agc ttc atc cag agc atc aaa gtg atc aac gcc atc Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile 450 455 460	1392
atc aag aag tac ggc ctg ccc aac gac atc att atc gag ctg gcc cgc Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg 465 470 475 480	1440
gag aag aac tcc aag gac gcc cag aaa atg atc aac gag atg cag aag	1488

Glu	Lys	Asn	Ser	Lys	Asp	Ala	Gln	Lys	Met	Ile	Asn	Glu	Met	Gln	Lys	
				485					490					495		
cgg	aac	cgg	cag	acc	aac	gag	cgg	atc	gag	gaa	atc	atc	cgg	acc	acc	1536
Arg	Asn	Arg	Gln	Thr	Asn	Glu	Arg	Ile	Glu	Glu	Ile	Ile	Arg	Thr	Thr	
			500					505					510			
ggc	aaa	gag	aac	gcc	aag	tac	ctg	atc	gag	aag	atc	aag	ctg	cac	gac	1584
Gly	Lys	Glu	Asn	Ala	Lys	Tyr	Leu	Ile	Glu	Lys	Ile	Lys	Leu	His	Asp	
		515					520					525				
atg	cag	gaa	ggc	aag	tgc	ctg	tac	agc	ctg	gaa	gcc	atc	cct	ctg	gaa	1632
Met	Gln	Glu	Gly	Lys	Cys	Leu	Tyr	Ser	Leu	Glu	Ala	Ile	Pro	Leu	Glu	
	530					535					540					
gat	ctg	ctg	aac	aac	ccc	ttc	aac	tat	gag	gtg	gac	cac	atc	atc	ccc	1680
Asp	Leu	Leu	Asn	Asn	Pro	Phe	Asn	Tyr	Glu	Val	Asp	His	Ile	Ile	Pro	
545					550					555					560	
aga	agc	gtg	tcc	ttc	gac	aac	agc	ttc	aac	aac	aag	gtg	ctc	gtg	aag	1728
Arg	Ser	Val	Ser	Phe	Asp	Asn	Ser	Phe	Asn	Asn	Lys	Val	Leu	Val	Lys	
				565					570					575		
cag	gaa	gaa	gcc	agc	aag	aag	ggc	aac	cgg	acc	cca	ttc	cag	tac	ctg	1776
Gln	Glu	Glu	Ala	Ser	Lys	Lys	Gly	Asn	Arg	Thr	Pro	Phe	Gln	Tyr	Leu	
			580					585					590			
agc	agc	agc	gac	agc	aag	atc	agc	tac	gaa	acc	ttc	aag	aag	cac	atc	1824
Ser	Ser	Ser	Asp	Ser	Lys	Ile	Ser	Tyr	Glu	Thr	Phe	Lys	Lys	His	Ile	
		595					600					605				
ctg	aat	ctg	gcc	aag	ggc	aag	ggc	aga	atc	agc	aag	acc	aag	aaa	gag	1872
Leu	Asn	Leu	Ala	Lys	Gly	Lys	Gly	Arg	Ile	Ser	Lys	Thr	Lys	Lys	Glu	
	610					615					620					
tat	ctg	ctg	gaa	gaa	cgg	gac	atc	aac	agg	ttc	tcc	gtg	cag	aaa	gac	1920
Tyr	Leu	Leu	Glu	Glu	Arg	Asp	Ile	Asn	Arg	Phe	Ser	Val	Gln	Lys	Asp	
625					630					635					640	
ttc	atc	aac	cgg	aac	ctg	gtg	gat	acc	aga	tac	gcc	acc	aga	ggc	ctg	1968
Phe	Ile	Asn	Arg	Asn	Leu	Val	Asp	Thr	Arg	Tyr	Ala	Thr	Arg	Gly	Leu	
				645					650					655		
atg	aac	ctg	ctg	cgg	agc	tac	ttc	aga	gtg	aac	aac	ctg	gac	gtg	aaa	2016
Met	Asn	Leu	Leu	Arg	Ser	Tyr	Phe	Arg	Val	Asn	Asn	Leu	Asp	Val	Lys	
			660				665						670			
gtg	aag	tcc	atc	aat	ggc	ggc	ttc	acc	agc	ttt	ctg	cgg	cgg	aag	tgg	2064
Val	Lys	Ser	Ile	Asn	Gly	Gly	Phe	Thr	Ser	Phe	Leu	Arg	Arg	Lys	Trp	
		675					680					685				
aag	ttt	aag	aaa	gag	cgg	aac	aag	ggg	tac	aag	cac	cac	gcc	gag	gac	2112
Lys	Phe	Lys	Lys	Glu	Arg	Asn	Lys	Gly	Tyr	Lys	His	His	Ala	Glu	Asp	
	690					695					700					
gcc	ctg	atc	att	gcc	aac	gcc	gat	ttc	atc	ttc	aaa	gag	tgg	aag	aaa	2160
Ala	Leu	Ile	Ile	Ala	Asn	Ala	Asp	Phe	Ile	Phe	Lys	Glu	Trp	Lys	Lys	
705					710					715					720	

tcc ggc ggc ggt tgc acc gag cag gag tac aaa gag atc ttc atc acc Ser Gly Gly Gly Ser Thr Glu Gln Glu Tyr Lys Glu Ile Phe Ile Thr 725 730 735	2208
ccc cac cag atc aag cac att aag gac ttc aag gac tac aag tac agc Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr Lys Tyr Ser 740 745 750	2256
cac cgg gtg gac aag aag cct aat aga gag ctg att aac gac acc ctg His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile Asn Asp Thr Leu 755 760 765	2304
tac tcc acc cgg aag gac gac aag ggc aac acc ctg atc gtg aac aat Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn Asn 770 775 780	2352
ctg aac ggc ctg tac gac aag gac aat gac aag ctg aaa aag ctg atc Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile 785 790 795 800	2400
aac aag agc ccc gaa aag ctg ctg atg tac cac cac gac ccc cag acc Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr 805 810 815	2448
tac cag aaa ctg aag ctg att atg gaa cag tac ggc gac gag aag aat Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly Asp Glu Lys Asn 820 825 830	2496
ccc ctg tac aag tac tac gag gaa acc ggg aac tac ctg acc aag tac Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu Thr Lys Tyr 835 840 845	2544
tcc aaa aag gac aac ggc ccc gtg atc aag aag att aag tat tac ggc Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys Tyr Tyr Gly 850 855 860	2592
aac aaa ctg aac gcc cat ctg gac atc acc gac gac tac ccc aac agc Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr Pro Asn Ser 865 870 875 880	2640
aga aac aag gtc gtg aag ctg tcc ctg aag ccc tac aga ttc gac gtg Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg Phe Asp Val 885 890 895	2688
tac ctg gac aat ggc gtg tac aag ttc gtg acc gtg aag aat ctg gat Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys Asn Leu Asp 900 905 910	2736
gtg atc aaa aaa gaa aac tac tac gaa gtg aat agc aag tgc tat gag Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys Cys Tyr Glu 915 920 925	2784
gaa gct aag aag ctg aag aag atc agc aac cag gcc gag ttt atc gcc Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu Phe Ile Ala 930 935 940	2832
tcc ttc tac aac aac gat ctg atc aag atc aac ggc gag ctg tat aga	2880

Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly Glu Leu Tyr Arg
 945 950 955 960
 gtg atc ggc gtg aac aac gac ctg ctg aac cgg atc gaa gtg aac atg 2928
 Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu Val Asn Met
 965 970 975
 atc gac atc acc tac cgc gag tac ctg gaa aac atg aac gac aag agg 2976
 Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn Asp Lys Arg
 980 985 990
 ccc ccc agg atc att aag aca atc gcc tcc aag acc cag agc att aag 3024
 Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr Gln Ser Ile Lys
 995 1000 1005
 aag tac agc aca gac att ctg ggc aac ctg tat gaa gtg aaa tct 3069
 Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr Glu Val Lys Ser
 1010 1015 1020
 aag aag cac cct cag atc atc aaa aag ggc taa 3102
 Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1025 1030

<210> 4
 <211> 1033
 <212> PRT
 <213> Staphylococcus aureus

<400> 4

Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
 1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
 20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
 35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
 50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
 65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
 85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
 100 105 110

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

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
130 135 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
145 150 155 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
225 230 235 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala

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

Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
690 695 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705 710 715 720

Ser Gly Gly Gly Ser Thr Glu Gln Glu Tyr Lys Glu Ile Phe Ile Thr
725 730 735

Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr Lys Tyr Ser
740 745 750

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

Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile Val Asn Asn
770 775 780

Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys Lys Leu Ile
785 790 795 800

Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp Pro Gln Thr

805										810				815			
Tyr	Gln	Lys	Leu	Lys	Leu	Ile	Met	Glu	Gln	Tyr	Gly	Asp	Glu	Lys	Asn		
			820					825					830				
Pro	Leu	Tyr	Lys	Tyr	Tyr	Glu	Glu	Thr	Gly	Asn	Tyr	Leu	Thr	Lys	Tyr		
		835					840					845					
Ser	Lys	Lys	Asp	Asn	Gly	Pro	Val	Ile	Lys	Lys	Ile	Lys	Tyr	Tyr	Gly		
	850					855					860						
Asn	Lys	Leu	Asn	Ala	His	Leu	Asp	Ile	Thr	Asp	Asp	Tyr	Pro	Asn	Ser		
865					870					875					880		
Arg	Asn	Lys	Val	Val	Lys	Leu	Ser	Leu	Lys	Pro	Tyr	Arg	Phe	Asp	Val		
				885					890					895			
Tyr	Leu	Asp	Asn	Gly	Val	Tyr	Lys	Phe	Val	Thr	Val	Lys	Asn	Leu	Asp		
			900					905					910				
Val	Ile	Lys	Lys	Glu	Asn	Tyr	Tyr	Glu	Val	Asn	Ser	Lys	Cys	Tyr	Glu		
		915					920					925					
Glu	Ala	Lys	Lys	Leu	Lys	Lys	Ile	Ser	Asn	Gln	Ala	Glu	Phe	Ile	Ala		
	930					935					940						
Ser	Phe	Tyr	Asn	Asn	Asp	Leu	Ile	Lys	Ile	Asn	Gly	Glu	Leu	Tyr	Arg		
945					950					955					960		
Val	Ile	Gly	Val	Asn	Asn	Asp	Leu	Leu	Asn	Arg	Ile	Glu	Val	Asn	Met		
				965					970					975			
Ile	Asp	Ile	Thr	Tyr	Arg	Glu	Tyr	Leu	Glu	Asn	Met	Asn	Asp	Lys	Arg		
			980					985					990				
Pro	Pro	Arg	Ile	Ile	Lys	Thr	Ile	Ala	Ser	Lys	Thr	Gln	Ser	Ile	Lys		
		995					1000					1005					
Lys	Tyr	Ser	Thr	Asp	Ile	Leu	Gly	Asn	Leu	Tyr	Glu	Val	Lys	Ser			
	1010					1015					1020						
Lys	Lys	His	Pro	Gln	Ile	Ile	Lys	Lys	Gly								
	1025					1030											

<210> 5
 <211> 3072
 <212> DNA
 <213> Staphylococcus aureus

<220>
 <221> CDS
 <222> (1).. (3072)

<400> 5	
atg aag cgg aac tac atc ctg ggc ctg gcc atc ggc atc acc agc gtg	48
Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val	
1 5 10 15	
ggc tac ggc atc atc gac tac gag aca cgg gac gtg atc gat gcc ggc	96
Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly	
20 25 30	
gtg cgg ctg ttc aaa gag gcc aac gtg gaa aac aac gag ggc agg cgg	144
Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg	
35 40 45	
agc aag aga ggc gcc aga agg ctg aag cgg cgg agg cgg cat aga atc	192
Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile	
50 55 60	
cag aga gtg aag aag ctg ctg ttc gac tac aac ctg ctg acc gac cac	240
Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His	
65 70 75 80	
agc gag ctg agc ggc atc aac ccc tac gag gcc aga gtg aag ggc ctg	288
Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu	
85 90 95	
agc cag aag ctg agc gag gaa gag ttc tct gcc gcc ctg ctg cac ctg	336
Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu	
100 105 110	
gcc aag aga aga ggc gtg cac aac gtg aac gag gtg gaa gag gac acc	384
Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr	
115 120 125	
ggc aac gag ctg tcc acc aaa gag cag atc agc cgg aac agc aag gcc	432
Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala	
130 135 140	
ctg gaa gag aaa tac gtg gcc gaa ctg cag ctg gaa cgg ctg aag aaa	480
Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys	
145 150 155 160	
gac ggc gaa gtg cgg ggc agc atc aac aga ttc aag acc agc gac tac	528
Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr	
165 170 175	
gtg aaa gaa gcc aaa cag ctg ctg aag gtg cag aag gcc tac cac cag	576
Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln	

180						185						190						
ctg Leu	gac Asp	cag Gln 195	agc Ser	ttc Phe	atc Ile	gac Asp	acc Thr 200	tac Tyr	atc Ile	gac Asp	ctg Leu	ctg Leu 205	gaa Glu	acc Thr	cgg Arg	624		
cgg Arg	acc Thr 210	tac Tyr	tat Tyr	gag Glu	gga Gly	cct Pro 215	ggc Gly	gag Glu	ggc Gly	agc Ser	ccc Pro 220	ttc Phe	ggc Gly	tgg Trp	aag Lys	672		
gac Asp 225	atc Ile	aaa Lys	gaa Glu	tgg Trp	tac Tyr 230	gag Glu	atg Met	ctg Leu	atg Met	ggc Gly 235	cac His	tgc Cys	acc Thr	tac Tyr	ttc Phe 240	720		
ccc Pro	gag Glu	gaa Glu	ctg Leu	cgg Arg 245	agc Ser	gtg Val	aag Lys	tac Tyr	gcc Ala 250	tac Tyr	aac Asn	gcc Ala	gac Asp	ctg Leu 255	tac Tyr	768		
aac Asn	gcc Ala	ctg Leu	aac Asn 260	gac Asp	ctg Leu	aac Asn	aat Asn	ctc Leu 265	gtg Val	atc Ile	acc Thr	agg Arg	gac Asp 270	gag Glu	aac Asn	816		
gag Glu	aag Lys	ctg Leu 275	gaa Glu	tat Tyr	tac Tyr	gag Glu	aag Lys 280	ttc Phe	cag Gln	atc Ile	atc Ile	gag Glu 285	aac Asn	gtg Val	ttc Phe	864		
aag Lys	cag Gln 290	aag Lys	aag Lys	aag Lys	ccc Pro	acc Thr 295	ctg Leu	aag Lys	cag Gln	atc Ile	gcc Ala 300	aaa Lys	gaa Glu	atc Ile	ctc Leu	912		
gtg Val 305	aac Asn	gaa Glu	gag Glu	gat Asp	att Ile 310	aag Lys	ggc Gly	tac Tyr	aga Arg	gtg Val 315	acc Thr	agc Ser	acc Thr	ggc Gly	aag Lys 320	960		
ccc Pro	gag Glu	ttc Phe	acc Thr	aac Asn 325	ctg Leu	aag Lys	gtg Val	tac Tyr	cac His 330	gac Asp	atc Ile	aag Lys	gac Asp	att Ile 335	acc Thr	1008		
gcc Ala	cgg Arg	aaa Lys	gag Glu 340	att Ile	att Ile	gag Glu	aac Asn	gcc Ala 345	gag Glu	ctg Leu	ctg Leu	gat Asp	cag Gln 350	att Ile	gcc Ala	1056		
aag Lys	atc Ile	ctg Leu 355	acc Thr	atc Ile	tac Tyr	cag Gln	agc Ser 360	agc Ser	gag Glu	gac Asp	atc Ile	cag Gln 365	gaa Glu	gaa Glu	ctg Leu	1104		
acc Thr	aat Asn 370	ctg Leu	aac Asn	tcc Ser	gag Glu	ctg Leu 375	acc Thr	cag Gln	gaa Glu	gag Glu	atc Ile 380	gag Glu	cag Gln	atc Ile	tot Ser	1152		
aat Asn 385	ctg Leu	aag Lys	ggc Gly	tat Tyr	acc Thr 390	ggc Gly	acc Thr	cac His	aac Asn	ctg Leu 395	agc Ser	ctg Leu	aag Lys	gcc Ala	atc Ile 400	1200		
aac Asn	ctg Leu	atc Ile	ctg Leu	gac Asp 405	gag Glu	ctg Leu	tgg Trp	cac His	acc Thr 410	aac Asn	gac Asp	aac Asn	cag Gln	atc Ile 415	gct Ala	1248		

atc ttc aac cgg ctg aag ctg gtg ccc aag aag gtg gac ctg tcc cag Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln 420 425 430	1296
cag aaa gag atc ccc acc acc ctg gtg gac gac ttc atc ctg agc ccc Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro 435 440 445	1344
gtc gtg aag aga agc ttc atc cag agc atc aaa gtg atc aac gcc atc Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile 450 455 460	1392
atc aag aag tac ggc ctg ccc aac gac atc att atc gag ctg gcc cgc Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg 465 470 475 480	1440
gag aag aac tcc aag gac gcc cag aaa atg atc aac gag atg cag aag Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys 485 490 495	1488
cgg aac cgg cag acc aac gag cgg atc gag gaa atc atc cgg acc acc Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr 500 505 510	1536
ggc aaa gag aac gcc aag tac ctg atc gag aag atc aag ctg cac gac Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp 515 520 525	1584
atg cag gaa ggc aag tgc ctg tac agc ctg gaa gcc atc cct ctg gaa Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu 530 535 540	1632
gat ctg ctg aac aac ccc ttc aac tat gag gtg gac cac atc atc ccc Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro 545 550 555 560	1680
aga agc gtg tcc ttc gac aac agc ttc aac aac aag gtg ctc gtg aag Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys 565 570 575	1728
cag gaa gaa gcc agc aag aag ggc aac cgg acc cca ttc cag tac ctg Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu 580 585 590	1776
agc agc agc gac agc aag atc agc tac gaa acc ttc aag aag cac atc Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile 595 600 605	1824
ctg aat ctg gcc aag ggc aag ggc aga atc agc aag acc aag aaa gag Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu 610 615 620	1872
tat ctg ctg gaa gaa cgg gac atc aac agg ttc tcc gtg cag aaa gac Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp 625 630 635 640	1920
ttc atc aac cgg aac ctg gtg gat acc aga tac gcc acc aga ggc ctg Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu 645 650 655 660	1968

645								650					655					
atg Met	aac Asn	ctg Leu	ctg Leu 660	cgg Arg	agc Ser	tac Tyr	ttc Phe	aga Arg 665	gtg Val	aac Asn	aac Asn	ctg Leu	gac Asp 670	gtg Val	aaa Lys	2016		
gtg Val	aag Lys	tcc Ser 675	atc Ile	aat Asn	ggc Gly	ggc Gly	ttc Phe 680	acc Thr	agc Ser	ttt Phe	ctg Leu	cgg Arg 685	cgg Arg	aag Lys	tgg Trp	2064		
aag Lys	ttt Phe 690	aag Lys	aaa Lys	gag Glu	cgg Arg	aac Asn 695	aag Lys	ggg Gly	tac Tyr	aag Lys	cac His 700	cac His	gcc Ala	gag Glu	gac Asp	2112		
gcc Ala 705	ctg Leu	atc Ile	att Ile	gcc Ala	aac Asn 710	gcc Ala	gat Asp	ttc Phe	atc Ile	ttc Phe 715	aaa Lys	gag Glu	tgg Trp	aag Lys	aaa Lys 720	2160		
tcc Ser	ggc Gly	ggc Gly	ggt Gly	tcc Ser 725	acc Thr	ccc Pro	cac His	cag Gln	atc Ile 730	aag Lys	cac His	att Ile	aag Lys	gac Asp 735	ttc Phe	2208		
aag Lys	gac Asp	tac Tyr	aag Lys 740	tac Tyr	agc Ser	cac His	cgg Arg	gtg Val 745	gac Asp	aag Lys	aag Lys	cct Pro	aat Asn 750	aga Arg	gag Glu	2256		
ctg Leu	att Ile	aac Asn 755	gac Asp	acc Thr	ctg Leu	tac Tyr	tcc Ser 760	acc Thr	cgg Arg	aag Lys	gac Asp	gac Asp 765	aag Lys	ggc Gly	aac Asn	2304		
acc Thr	ctg Leu 770	atc Ile	gtg Val	aac Asn	aat Asn	ctg Leu 775	aac Asn	ggc Gly	ctg Leu	tac Tyr	gac Asp 780	aag Lys	gac Asp	aat Asn	gac Asp	2352		
aag Lys 785	ctg Leu	aaa Lys	aag Lys	ctg Leu	atc Ile 790	aac Asn	aag Lys	agc Ser	ccc Pro	gaa Glu 795	aag Lys	ctg Leu	ctg Leu	atg Met	tac Tyr 800	2400		
cac His	cac His	gac Asp	ccc Pro	cag Gln 805	acc Thr	tac Tyr	cag Gln	aaa Lys	ctg Leu 810	aag Lys	ctg Leu	att Ile	atg Met	gaa Glu 815	cag Gln	2448		
tac Tyr	ggc Gly	gac Asp	gag Glu 820	aag Lys	aat Asn	ccc Pro	ctg Leu	tac Tyr 825	aag Lys	tac Tyr	tac Tyr	gag Glu	gaa Glu 830	acc Thr	ggg Gly	2496		
aac Asn	tac Tyr	ctg Leu 835	acc Thr	aag Lys	tac Tyr	tcc Ser	aaa Lys 840	aag Lys	gac Asp	aac Asn	ggc Gly	ccc Pro 845	gtg Val	atc Ile	aag Lys	2544		
aag Lys	att Ile 850	aag Lys	tat Tyr	tac Tyr	ggc Gly	aac Asn 855	aaa Lys	ctg Leu	aac Asn	gcc Ala	cat His 860	ctg Leu	gac Asp	atc Ile	acc Thr	2592		
gac Asp 865	gac Asp	tac Tyr	ccc Pro	aac Asn	agc Ser 870	aga Arg	aac Asn	aag Lys	gtc Val	gtg Val 875	aag Lys	ctg Leu	tcc Ser	ctg Leu	aag Lys 880	2640		

ccc tac aga ttc gac gtg tac ctg gac aat ggc gtg tac aag ttc gtg 2688
Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val
885 890 895

acc gtg aag aat ctg gat gtg atc aaa aaa gaa aac tac tac gaa gtg 2736
Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val
900 905 910

aat agc aag tgc tat gag gaa gct aag aag ctg aag aag atc agc aac 2784
Asn Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn
915 920 925

cag gcc gag ttt atc gcc tcc ttc tac aac aac gat ctg atc aag atc 2832
Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile
930 935 940

aac ggc gag ctg tat aga gtg atc ggc gtg aac aac gac ctg ctg aac 2880
Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn
945 950 955 960

cgg atc gaa gtg aac atg atc gac atc acc tac cgc gag tac ctg gaa 2928
Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu
965 970 975

aac atg aac gac aag agg ccc ccc agg atc att aag aca atc gcc tcc 2976
Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser
980 985 990

aag acc cag agc att aag aag tac agc aca gac att ctg ggc aac ctg 3024
Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
995 1000 1005

tat gaa gtg aaa tct aag aag cac cct cag atc atc aaa aag ggc 3069
Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1010 1015 1020

taa 3072

<210> 6
<211> 1023
<212> PRT
<213> Staphylococcus aureus

<400> 6

Met Lys Arg Asn Tyr Ile Leu Gly Leu Ala Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80

Ser Glu Leu Ser Gly Ile Asn Pro Tyr Glu Ala Arg Val Lys Gly Leu
85 90 95

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
100 105 110

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

Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
130 135 140

Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
145 150 155 160

Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
165 170 175

Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
180 185 190

Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
195 200 205

Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
210 215 220

Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
225 230 235 240

Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
245 250 255

Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
260 265 270

Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
275 280 285

Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
290 295 300

Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
305 310 315 320

Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
325 330 335

Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala
340 345 350

Lys Ile Leu Thr Ile Tyr Gln Ser Ser Glu Asp Ile Gln Glu Glu Leu
355 360 365

Thr Asn Leu Asn Ser Glu Leu Thr Gln Glu Glu Ile Glu Gln Ile Ser
370 375 380

Asn Leu Lys Gly Tyr Thr Gly Thr His Asn Leu Ser Leu Lys Ala Ile
385 390 395 400

Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430

Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445

Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
450 455 460

Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480

Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495

Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510

Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525

Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540

Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
545 550 555 560

Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575

Gln Glu Glu Ala Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605

Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620

Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640

Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655

Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
660 665 670

Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
675 680 685

Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
690 695 700

Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
705 710 715 720

Ser Gly Gly Gly Ser Thr Pro His Gln Ile Lys His Ile Lys Asp Phe
725 730 735

Lys Asp Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu
740 745 750

Leu Ile Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn
755 760 765

Thr Leu Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp
770 775 780

Lys Leu Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr
785 790 795 800

His His Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln
805 810 815

Tyr Gly Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly
820 825 830

Asn Tyr Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys
835 840 845

Lys Ile Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr
850 855 860

Asp Asp Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys
865 870 875 880

Pro Tyr Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val
885 890 895

Thr Val Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val
900 905 910

Asn Ser Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn
915 920 925

Gln Ala Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile
930 935 940

Asn Gly Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn
945 950 955 960

Arg Ile Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu
965 970 975

Asn Met Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser
 980 985 990

Lys Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
 995 1000 1005

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1010 1015 1020

<210> 7
 <211> 5
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Linker

<400> 7

Ser Gly Gly Gly Ser
 1 5

<210> 8
 <211> 6
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Linker

<400> 8

Gly Gly Ser Gly Gly Ser
 1 5

<210> 9
 <211> 8
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Linker

<400> 9

Ser Gly Ser Gly Ser Gly Ser Gly
 1 5

<210> 10
 <211> 9
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Linker
 <400> 10
 Ser Gly Ser Gly Ser Gly Ser Gly Ser
 1 5

<210> 11
 <211> 20
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> crRNA (C1)

<400> 11
 acggaggcua agcgucgcaa 20

<210> 12
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> crRNA (A2)

<400> 12
 ggagccacag uucuuaccag g 21

<210> 13
 <211> 21
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> crRNA (A3)

<400> 13
 cucuacccuu gaggucucga g 21

<210> 14
 <211> 83
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> tracrRNA

<400> 14
 guuuuaguac ucuggaaaca gaauacuua aaacaaggca aaaugccgug uuuaucacgu 60

caacuuguug gcgagauuuu uuu 83

CLAIMS

1. A protein having a binding ability to guide RNA and consisting of a sequence comprising an amino acid sequence
5 wherein a continuous deletion region is present between the 721-position and the 755-position in the amino acid sequence shown in SEQ ID NO: 2,
wherein amino acids adjacent to each of the deletion region are linked by a linker consisting of 3 to 10 amino acid residues.
- 10 2. The protein according to claim 1, wherein the linker is a 5 - 9 amino acid length linker composed of glycine (G) and serine (S).
- 15 3. The protein according to claim 2, wherein the linker is selected from the following:
-SGGGS-
-GGSGGS-
-SGSGSGSG-
20 -SGSGSGSGS-.
4. The protein according to any one of claims 1 to 3, wherein the deletion region is a region of the 721-position to the 745-position.
- 25 5. The protein according to claim 4, wherein the protein is shown in SEQ ID NO: 4.
6. The protein according to any one of claims 1 to 3, wherein
30 the deletion region is a region of the 721-position to the 755-position.
7. The protein according to claim 6, wherein the protein is shown in SEQ ID NO: 6.

35

8. The protein according to any one of claims 1 to 7, wherein glutamic acid (E) at the 45-position and/or the 163-position are/is substituted with other amino acid(s).

5 9. The protein according to claim 8, wherein said other amino acid is a basic amino acid.

10. The protein according to claim 9, wherein the basic amino acid is lysine (K).

10

11. The protein according to any one of claims 1 to 10, having identity of 80% or more at a site other than the mutated and/or deleted positions in the SEQ ID NO: 2.

15 12. The protein according to any one of claims 1 to 10, wherein one to several amino acids are substituted, deleted, inserted and/or added at a site other than the mutated and/or deleted positions in the SEQ ID NO: 2.

20 13. The protein according to any one of claims 1 to 12, wherein a transcriptional regulator protein or domain is linked.

14. The protein according to claim 13, wherein the transcriptional regulator is a transcriptional activator.

25

15. The protein according to claim 13, wherein the transcriptional regulator is a transcriptional silencer or a transcriptional inhibitor.

30 16. A nucleic acid encoding the protein according to any one of claims 1 to 15.

17. A protein-RNA complex provided with the protein according to any one of claims 1 to 16 and a guide RNA comprising a
35 polynucleotide composed of a base sequence complementary to a

base sequence located 1 to 20 to 24 bases upstream from a proto-spacer adjacent motif (PAM) sequence in a target double-stranded polynucleotide.

5 18. A method for site-specifically modifying a target double-stranded polynucleotide, including
a step of mixing and incubating a target double-stranded polynucleotide, a protein and a guide RNA, and
a step of having the aforementioned protein modify the
10 aforementioned target double-stranded polynucleotide at a binding site located upstream of a PAM sequence; wherein,
the aforementioned protein is the protein according to any one of claims 1 to 15, and
the aforementioned guide RNA contains a polynucleotide
15 composed of a base sequence complementary to a base sequence located 1 to 20 to 24 bases upstream from the aforementioned PAM sequence in the aforementioned target double-stranded polynucleotide.

20 19. A method for increasing expression of a target gene in a cell, comprising expressing the protein according to claim 14 and one or plural guide RNAs for the aforementioned target gene in the aforementioned cell.

25 20. A method for decreasing expression of a target gene in a cell, comprising expressing the protein according to claim 15 and one or plural guide RNAs for the aforementioned target gene in the aforementioned cell.

30 21. The method according to claim 19 or 20, wherein the cell is a eukaryotic cell.

22. The method according to claim 19 or 20, wherein the cell is a yeast cell, a plant cell or an animal cell.