Claims

1. A protein having an amino acid sequence as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19 or an ortholog, homolog, variant or functional fragment thereof; wherein, the ortholog, homolog, variant or functional fragment substantially retains the biological function of the sequence from which it is derived;

for example, the ortholog, homolog, or variant has at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity compared to the sequence from which it is derived;

for example, the ortholog, homolog, or variant has at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity compared with the sequence as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19, and substantially retains the biological functions of the sequence from which it is derived;

for example, the protein is an effector protein in the CRISPR/Cas system.

2. The protein of claim 1, which comprises a sequence selected from the following, or consists of a sequence selected from the following:

(i) a sequence as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19;

(ii) compared with the sequence as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19, a sequence having one or more amino acid substitutions, deletions or additions (for example, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 amino acid substitutions, deletions or additions); or

(iii) a sequence having at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity with the sequence

as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19;

for example, the protein has an amino acid sequence as shown in any one of SEQ ID NOs: 17, 2, 20, 16, 19.

3. A conjugate comprising the protein of claim 1 or 2 and a modified portion;

for example, the modified portion is selected from an additional protein or polypeptide, a detectable label, and any combinations thereof;

for example, the modified portion is optionally connected to the N-terminus or C-terminus of the protein through a linker;

for example, the modified portion is fused to the N-terminus or C-terminus of the protein;

for example, the additional protein or polypeptide is selected from an epitope tag, a reporter gene sequence, a nuclear localization signal (NLS) sequence, a targeting moiety, a transcription activation domain (such as, VP64), a transcription repression domain (for example, KRAB domain or SID domain), a nuclease domain (for example, Fok1), a domain having an activity selected from: nucleotide deaminase, methylase activity, demethylase, transcription activation activity, transcription inhibition activity, transcription release factor activity, histone modification activity, nuclease activity, single-stranded RNA cleavage activity, double-stranded RNA cleavage activity, single-stranded DNA cleavage activity, double-stranded DNA cleavage activity and nucleic acid binding activity; and any combinations thereof;

for example, the conjugate comprises an epitope tag;

for example, the conjugate comprises an NLS sequence;

for example, the NLS sequence is shown in SEQ ID NO: 81;

for example, the NLS sequence is located at, near or close to the end of the protein (e.g., N-terminal or C-terminal).

4. A fusion protein comprising the protein of claim 1 or 2 and an additional protein or polypeptide;

for example, the additional protein or polypeptide is optionally linked to the N-terminus or C-terminus of the protein through a linker;

for example, the additional protein or polypeptide is selected from an epitope tag, a reporter gene sequence, a nuclear localization signal (NLS) sequence, a targeting moiety, a transcription activation domain (such as, VP64), a transcription repression domain (for example, KRAB domain or SID domain), a nuclease domain (for example, Fok1), a domain having an activity selected from: a nucleotide deaminase, methylase activity, a demethylase, transcription activation activity, transcription inhibition activity, transcription release factor activity, histone modification activity, nuclease activity, single-stranded RNA cleavage activity, double-stranded RNA cleavage activity, single-stranded DNA cleavage activity, double-stranded DNA cleavage activity and nucleic acid binding activity ; and any combinations thereof;

for example, the fusion protein comprises an epitope tag;

for example, the fusion protein comprises an NLS sequence;

for example, the NLS sequence is shown in SEQ ID NO: 81;

for example, the NLS sequence is located at, near, or close to the end of the protein (for example, the N-terminus or the C-terminus);

for example, the fusion protein has an amino acid sequence selected from: SEQ ID NOs: 98, 83, 101, 97, 100.

5. An isolated nucleic acid molecule comprising a sequence selected from the following or consisting of a sequence selected from the following:

(i) a sequence as shown in any one of SEQ ID NOs: 57, 42, 60, 56, 59;

(ii) compared with the sequence as shown in any one of SEQ ID NOs:

57, 42, 60, 56, 59, a sequence having one or more base substitutions, deletions or additions (for example, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 base substitutions, deletions or additions);

(iii) a sequence having at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% sequence identity with the sequence as shown in any one of SEQ ID NO: 57, 42, 60, 56, 59;

(iv) a sequence that hybridizes to the sequence as described in any one of (i) to (iii) under stringent conditions; or

(v) a complementary sequence of the sequence as described in any one of (i)-(iv);

in addition, the sequence as described in any one of (i)-(v) substantially retains the biological function of the sequence from which it is derived;

for example, the nucleic acid molecule contains one or more stem loops or optimized secondary structures;

for example, the sequence described in any one of (i)-(v) retains the secondary structure of the sequence from which it is derived;

for example, the nucleic acid molecule comprises a sequence selected from the following, or consists of a sequence selected from the following:

(a) a nucleotide sequence as shown in any one of SEQ ID NOs: 57, 42, 60, 56, 59;

(b) a sequence that hybridizes to the sequence as described in (a) under stringent conditions; or

(c) a complementary sequence of the sequence as described in (a);

for example, the isolated nucleic acid molecule is RNA;

for example, the isolated nucleic acid molecule is a direct repeat sequence in the CRISPR/Cas system.

6. A complex comprising:

(i) a protein component, which is selected from: the protein of claim 1 or 2, the conjugate of claim 3 or the fusion protein of claim 4, and any combinations thereof; and

(ii) a nucleic acid component, which comprises the isolated nucleic acid molecule of claim 5 and a guide sequence capable of hybridizing to the target sequence from the 5' to 3',

wherein the protein component and the nucleic acid component combine with each other to form a complex;

for example, the guide sequence is attached to the 3' end of the nucleic acid molecule;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the nucleic acid component is a guide RNA in the CRISPR/Cas system;

for example, the nucleic acid molecule is RNA;

for example, the complex does not comprise trans-acting crRNA (tracrRNA).

7. The complex of claim 6, comprising:

(i) a protein component selected from: the protein of claim 2, a conjugate or fusion protein comprising the protein; and

(ii) a nucleic acid component, which comprises the isolated nucleic acid molecule of claim 5 and the guide sequence.

8. An isolated nucleic acid molecule comprising:

(i) a nucleotide sequence encoding the protein of claim 1 or 2, or the conjugate of claim 3, or the fusion protein of claim 4;

(ii) a nucleotide sequence encoding the isolated nucleic acid molecule of claim 5; and/or,

(iii) a nucleotide sequence containing (i) and (ii);

for example, the nucleotide sequence described in any one of (i) to (iii) is codon optimized for expression in a prokaryotic cell or eukaryotic cell.

9. A vector comprising the isolated nucleic acid molecule of claim 8.

10. A host cell comprising the isolated nucleic acid molecule of claim 8 or the vector of claim 9.

11. A composition comprising:

(i) a first component, which is selected from: the protein of claim 1 or 2, the conjugate of claim 3, the fusion protein of claim 4, a nucleotide sequence encoding the protein or fusion protein, and any combinations thereof; and

(ii) a second component, which is a nucleotide sequence containing a guide RNA, or a nucleotide sequence encoding the nucleotide sequence containing a guide RNA;

wherein the guide RNA includes a direct repeat sequence and a guide sequence from the 5' to 3', and the guide sequence can hybridize with the target sequence;

the guide RNA can form a complex with the protein, conjugate or fusion protein as described in (i);

for example, the direct repeat sequence is an isolated nucleic acid molecule as defined in claim 5;

for example, the guide sequence is connected to the 3' end of the direct repeat sequence;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the composition does not contain trans-acting crRNA (tracrRNA);

for example, the composition is non-naturally occurring or modified;

for example, at least one component in the composition is non-naturally occurring or modified;

for example, the first component is non-naturally occurring or modified; and/or, the second component is non-naturally occurring or modified.

12. The composition of claim 11, wherein:

the first component is selected from: the protein of claim 2, or a conjugate or fusion protein comprising the protein, or a nucleotide sequence encoding the protein or fusion protein, and any combinations thereof;

the direct repeat sequence is an isolated nucleic acid molecule as defined in claim 5;

preferably, when the target sequence is DNA, the target sequence is located at the 3' end of the original spacer sequence adjacent motif (PAM), and the PAM has the sequence shown by 5'-ATG, 5'-TTN, or 5'-KTR.

13. A composition comprising one or more vectors, the one or more vectors comprising:

(i) a first nucleic acid, which is a nucleotide sequence encoding the protein of claim 1 or 2, or the fusion protein of claim 4; optionally, the first nucleic acid is operably linked to a first regulatory element; and

(ii) a second nucleic acid, which encodes a nucleotide sequence comprising a guide RNA; optionally the second nucleic acid is operably linked to a second regulatory element;

wherein:

the first nucleic acid and the second nucleic acid are present on the same or different vectors;

the guide RNA comprises a direct repeat sequence and a guide sequence from the 5' to 3', and the guide sequence can hybridize with the target sequence;

the guide RNA can form a complex with the effector protein or fusion protein as described in (i);

for example, the direct repeat sequence is an isolated nucleic acid molecule as defined in claim 5;

for example, the guide sequence is connected to the 3' end of the direct repeat sequence;

for example, the guide sequence comprises the complementary sequence of the target sequence;

for example, the composition does not contain trans-acting crRNA (tracrRNA);

for example, the composition is non-naturally occurring or modified;

for example, at least one component in the composition is non-naturally occurring or modified;

for example, the first regulatory element is a promoter, such as an inducible promoter;

for example, the second regulatory element is a promoter, such as an inducible promoter.

14. The composition of claim 13, wherein:

the first nucleic acid is a nucleotide sequence encoding the protein of claim 3 or a fusion protein containing the protein;

the direct repeat sequence is an isolated nucleic acid molecule as defined in claim 5;

preferably, when the target sequence is DNA, the target sequence is located at the 3' end of the original spacer sequence adjacent motif (PAM), and the PAM has the sequence shown by 5'-ATG, 5'-TTN, or 5'-KTR.

15. The composition of any one of claims 11-14, wherein when the target sequence is RNA, the target RNA sequence does not have PAM domain restrictions, and/or

wherein the target sequence is a DNA or RNA sequence derived from a prokaryotic cell or a eukaryotic cell; or the target sequence is a non-naturally occurring DNA or RNA sequence.

16. The composition of any one of claims 11-15, wherein the target sequence is present in a cell;

for example, the target sequence is present in the cell nucleus or in the cytoplasm (e.g., organelles);

for example, the cell is a eukaryotic cell;

for example, the cell is a prokaryotic cell.

17. The composition of any one of claims 11-16, wherein the protein is linked to one or more NLS sequences, or the conjugate or fusion protein comprises one or more NLS sequences;

for example, the NLS sequence is linked to the N-terminus or C-terminus of the protein;

for example, the NLS sequence is fused to the N-terminus or C-terminus of the protein.

18. A kit comprising one or more components selected from the group consisting of: the protein of claim 1 or 2, the conjugate of claim 3, the fusion protein of claim 4, the isolated nucleic acid molecule of claim 5, the complex of claim 6 or 7, the isolated nucleic acid molecule of claim 8, the vector of claim 9, the composition of any one of claims 11-17;

for example, the kit comprises the composition of claim 11 or 12, and instructions for using the composition;

for example, the kit comprises the composition of any one of claims 13-17, and instructions for using the composition.

19. A delivery composition comprising a delivery vehicle and one or

more selected from the group consisting of: the protein of claim 1 or 2, the conjugate of claim 3, the fusion protein of claim 4, the isolated nucleic acid molecule of claim 5, the complex of claim 6 or 7, the isolated nucleic acid molecule of claim 8, the vector of claim 9, the composition of any one of claims 11-17;

for example, the delivery vehicle is a particle;

for example, the delivery vehicle is selected from a lipid particle, sugar particle, metal particle, protein particle, liposome, exosome, microvesicle, gene gun, or viral vector (e.g., replication defective retrovirus, lentivirus, adenovirus or adeno-associated virus).

20. A method for modifying a target gene, comprising: contacting the complex of claim 6 or 7, or the composition of any one of claims 11-17 with the target gene, or delivering that to a cell containing the target gene; the target sequence is present in the target gene;

for example, the target gene is present in the cell;

wherein the cell is a prokaryotic cell or a eukaryotic cell or a plant cell;

for example, the target gene is present in a nucleic acid molecule (e.g., a plasmid) *in vitro*;

for example, the modification refers to a break in the target sequence, such as a double-strand break in DNA or a single-strand break in RNA;

for example, the modification further includes inserting an exogenous nucleic acid into the break.

21. A method for altering the expression of a gene product, comprising: contacting the complex of claim 6 or 7, or the composition of any one of claims 11-17 with a nucleic acid molecule encoding the gene product, or delivering that to a cell containing the nucleic acid molecule, the target sequence is present in the nucleic acid molecule;

for example, the nucleic acid molecule is present in the cell;

wherein the cell is a prokaryotic cell or a eukaryotic cell or a plant cell;

for example, the nucleic acid molecule is present in a nucleic acid molecule (e.g., a plasmid) *in vitro*;

for example, the expression of the gene product is altered (e.g., enhanced or decreased);

for example, the gene product is a protein.

22. The method of claim 20 or 21, wherein the protein, conjugate, fusion protein, isolated nucleic acid molecule, complex, vector or composition is contained in a delivery vehicle;

for example, the delivery vehicle is selected from a lipid particle, sugar particle, metal particle, protein particle, liposome, exosome, viral vector (such as replication-defective retrovirus, lentivirus, adenovirus or adeno-associated virus).

23. The method of any one of claims 20-22, which is used to change one or more target sequences in a target gene or a nucleic acid molecule encoding a target gene product to modify a cell, cell line, or organism.

24. A cell or its progeny obtained by the method of any one of claims 20-23, wherein the cell contains a modification that is not present in its wild type.

25. The cell product of the cell or its progeny of claim 24.

26. An *in vitro*, isolated cell or cell line or its progeny, the cell or cell line or its progeny comprises: the protein of claim 1 or 2, the conjugate of claim 3, the fusion protein of claim 4, the isolated nucleic acid molecule of claim 5, the complex of claim 6 or 7, the isolated nucleic acid molecule of claim 8, the vector of claim 9, the composition of any one of claims 11-17;

for example, the cell or cell line or its progeny comprises: the complex of claim 7, the composition of claim 12, or the composition of claim 14;

for example, the cell is a eukaryotic cell;

for example, the cell is an animal cell (for example, a mammalian cell, such as a human cell) or a plant cell;

for example, the cell is a stem cell or stem cell line.

27. Use of the protein of claim 1 or 2, the conjugate of claim 3, the fusion protein of claim 4, the isolated nucleic acid molecule of claim 5, the complex of claim 6 or 7, the isolated nucleic acid molecule of claim 8, the vector of claim 9, the composition of any one of claims 11-17, or the kit of claim 18 in the preparation of a formulation for:

(i) the *in vitro* gene or genome editing;

(ii) the detection of an isolated single-stranded DNA;

(iii) editing the target sequence in the target locus to modify a biological or non-human organism;

(iv) the treatment of the disease caused by defects in the target sequence in the target locus or

nucleic acid editing.