

Claims

[Claim 1]

A protein comprising at least two polypeptides of any one of the following polypeptides:

(A-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 9, or a PPR motif consisting of the sequence of SEQ ID NO: 9 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with tyrosine, substitution of the amino acid at position 15 with lysine, substitution of the amino acid at position 16 with leucine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 18 with aspartic acid, and substitution of the amino acid at position 28 with glutamic acid, or a PPR motif consisting of the sequence of SEQ ID NO: 401 or a PPR motif consisting of the sequence of SEQ ID NO: 401 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with tyrosine, substitution of the amino acid at position 16 with leucine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 18 with aspartic acid, and substitution of the amino acid at position 28 with glutamic acid;

(A-2') a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 9 or 401 having a substitution, deletion, or addition of 1 to 20 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 19, 26, 30, 33, and 34, and having an adenine-binding property, excluding a polypeptide in group A below;

(A-3') a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 42% to the sequence of SEQ ID NO: 9 or 401, provided that the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 19, 26, 30, 33, and 34 are identical, and having an adenine-binding property, excluding a polypeptide in group A below;

(C-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 10, or a PPR motif consisting of the sequence of SEQ ID NO: 10 having a substitution of amino acid selected from the group consisting of substitution of the amino acid at position 2 with serine, substitution of the amino acid at position 5 with isoleucine, substitution of the amino acid at position 7 with leucine, substitution of the amino acid at position 8 with lysine, substitution of the amino acid at position 10 with phenylalanine or tyrosine, substitution of the amino acid at position 15 with arginine, substitution of the amino acid at position 22 with valine, substitution of the amino acid

at position 24 with arginine, substitution of the amino acid at position 27 with leucine, and substitution of the amino acid at position 29 with arginine;

(C-2') a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 10 having a substitution, deletion, or addition of 1 to 25 amino acids other than the amino acids at positions 1, 3, 4, 14, 18, 19, 26, 30, 33, and 34, and having a cytosine-binding property, excluding a polypeptide in group C below;

(C-3') a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 25% to the sequence of SEQ ID NO: 10, provided that the amino acids at positions 1, 3, 4, 14, 18, 19, 26, 30, 33, and 34 are identical, and having a cytosine-binding property, excluding a polypeptide in group C below;

(G-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 11, or a PPR motif consisting of the sequence of SEQ ID NO: 11 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with phenylalanine, substitution of the amino acid at position 15 with aspartic acid, substitution of the amino acid at position 27 with valine, substitution of the amino acid at position 28 with serine, and substitution of the amino acid at position 35 with isoleucine;

(G-2') a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 11 having a substitution, deletion, or addition of 1 to 21 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 14, 18, 19, 26, 30, 33, and 34, and having a guanine-binding property, excluding a polypeptide in group G below;

(G-3') a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 40% to the sequence of SEQ ID NO: 11, provided that the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 14, 18, 19, 26, 30, 33, and 34 are identical, and having a guanine-binding property, excluding a polypeptide in group G below;

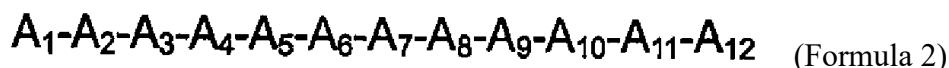
(U-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 12, or a PPR motif consisting of the sequence of SEQ ID NO: 12 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with phenylalanine, substitution of the amino acid at position 13 with serine, substitution of the amino acid at position 15 with lysine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 20 with leucine, substitution of the amino acid at position 21 with lysine, substitution of the amino acid at position 23 with phenylalanine, substitution of the amino acid at position 24 with aspartic acid, substitution of the amino acid at position 27 with lysine, substitution of the amino acid at position 28 with lysine, substitution of the amino acid at position 29

with arginine, and substitution of the amino acid at position 31 with leucine;
 (U-2') a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 12 having a substitution, deletion, or addition of 1 to 22 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 11, 12, 14, 19, 26, 30, 33, and 34, and having a uracil-binding property, excluding a polypeptide in group U below; and
 (U-3') a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 37% to the sequence of SEQ ID NO: 12, provided that the amino acids at positions 1, 2, 3, 4, 6, 11, 12, 14, 19, 26, 30, 33, and 34 are identical, and having a uracil-binding property, excluding a polypeptide in group U below.

(Helix A)-X-(Helix B)-L (Formula 1)

In the formula:

Helix A is a moiety of 12-amino acid length capable of forming an α -helix structure, and is represented by the formula 2;



wherein, in the formula 2, A₁ to A₁₂ independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an α -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L_{iii} to L_{vii} may not exist.

Group A:

- a polypeptide consisting of the sequence at position 545-579 of SEQ ID NO: 462,
- a polypeptide consisting of the sequence at position 409-443 of SEQ ID NO: 464,
- a polypeptide consisting of the sequence at position 431-465 of SEQ ID NO: 465,
- a polypeptide consisting of the sequence at position 437-471 of SEQ ID NO: 465

Group C:

- a polypeptide consisting of the sequence at position 318-352 of SEQ ID NO: 461

Group G:

a polypeptide consisting of the sequence at position 869-903 of SEQ ID NO: 463,
a polypeptide consisting of the sequence at position 374-408 of SEQ ID NO: 464,
a polypeptide consisting of the sequence at position 396-430 of SEQ ID NO: 465,
a polypeptide consisting of the sequence at position 402-436 of SEQ ID NO: 466

Group U:

a polypeptide consisting of the sequence at position 353-387 of SEQ ID NO: 461,
a polypeptide consisting of the sequence at position 475-509 of SEQ ID NO: 462,
a polypeptide consisting of the sequence at position 380-414 of SEQ ID NO: 463

[Claim 2]

The protein according to claim 1, wherein the polypeptides are any of followings:

the polypeptide of (A-1);

the polypeptide of (A-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 9 or 401 having a substitution, deletion, or addition of 1 to 9 amino acids;

the polypeptide of (A-3'), wherein the polypeptide having a sequence identity of at least 70% to the sequence of SEQ ID NO: 9 or 401;

the polypeptide of (C-1);

the polypeptide of (C-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 10 having a substitution, deletion, or addition of 1 to 9 amino acids;

the polypeptide of (CF-3'), wherein the polypeptide having a sequence identity of at least 70% to the sequence of SEQ ID NO: 10;

the polypeptide of (G-1);

the polypeptide of (G-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 11 having a substitution, deletion, or addition of 1 to 9 amino acids;

the polypeptide of (G-3'), wherein the polypeptide having a sequence identity of at least 70% to the sequence of SEQ ID NO: 11;

the polypeptide of (U-1);

the polypeptide of (U-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 12 having a substitution, deletion, or addition of 1 to 9 amino acids;

the polypeptide of (U-3'), wherein the polypeptide having a sequence identity of at least 70% to the sequence of SEQ ID NO: 12.

[Claim 3]

The protein according to claim 1 or 2, wherein the polypeptides are any of

followings:

the polypeptide of (A-1);

the polypeptide of (A-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 9 or 401 having a substitution, deletion, or addition of 1 to 4 amino acids;

the polypeptide of (A-3'), wherein the polypeptide having a sequence identity of at least 90% to the sequence of SEQ ID NO: 9 or 401;

the polypeptide of (C-1);

the polypeptide of (C-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 10 having a substitution, deletion, or addition of 1 to 4 amino acids;

the polypeptide of (CF-3'), wherein the polypeptide having a sequence identity of at least 90% to the sequence of SEQ ID NO: 10;

the polypeptide of (G-1);

the polypeptide of (G-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 11 having a substitution, deletion, or addition of 1 to 4 amino acids;

the polypeptide of (G-3'), wherein the polypeptide having a sequence identity of at least 90% to the sequence of SEQ ID NO: 11;

the polypeptide of (U-1);

the polypeptide of (U-2'), wherein the polypeptide consisting of the sequence of SEQ ID NO: 12 having a substitution, deletion, or addition of 1 to 4 amino acids;

the polypeptide of (U-3'), wherein the polypeptide having a sequence identity of at least 90% to the sequence of SEQ ID NO: 12.

[Claim 4]

Use of the polypeptide defined in any one of claims 1 to 3 for preparation of a PPR protein, or use of the polypeptide selected from the following, for preparation of a PPR protein of which target RNA has a length of 15 bases or longer.

(A-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 9, or a PPR motif consisting of the sequence of SEQ ID NO: 9 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with tyrosine, substitution of the amino acid at position 15 with lysine, substitution of the amino acid at position 16 with leucine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 18 with aspartic acid, and substitution of the amino acid at position 28 with glutamic acid, or a PPR motif consisting of the sequence of SEQ ID NO: 401 or a PPR motif consisting of the sequence of SEQ ID NO: 401 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with tyrosine, substitution of the amino acid at position 16

with leucine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 18 with aspartic acid, and substitution of the amino acid at position 28 with glutamic acid;

(A-2) a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 9 or 401 having a substitution, deletion, or addition of 1 to 20 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 19, 26, 30, 33, and 34, and having an adenine-binding property;

(A-3) a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 42% to the sequence of SEQ ID NO: 9 or 401, provided that the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 19, 26, 30, 33, and 34 are identical, and having an adenine-binding property;

(C-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 10, or a PPR motif consisting of the sequence of SEQ ID NO: 10 having a substitution of amino acid selected from the group consisting of substitution of the amino acid at position 2 with serine, substitution of the amino acid at position 5 with isoleucine, substitution of the amino acid at position 7 with leucine, substitution of the amino acid at position 8 with lysine, substitution of the amino acid at position 10 with phenylalanine or tyrosine, substitution of the amino acid at position 15 with arginine, substitution of the amino acid at position 22 with valine, substitution of the amino acid at position 24 with arginine, substitution of the amino acid at position 27 with leucine, and substitution of the amino acid at position 29 with arginine;

(C-2) a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 10 having a substitution, deletion, or addition of 1 to 25 amino acids other than the amino acids at positions 1, 3, 4, 14, 18, 19, 26, 30, 33, and 34, and having a cytosine-binding property;

(C-3) a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 25% to the sequence of SEQ ID NO: 10, provided that the amino acids at positions 1, 3, 4, 14, 18, 19, 26, 30, 33, and 34 are identical, and having a cytosine-binding property, ;

(G-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 11, or a PPR motif consisting of the sequence of SEQ ID NO: 11 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with phenylalanine, substitution of the amino acid at position 15 with aspartic acid, substitution of the amino acid at position 27 with valine, substitution of the amino acid at position 28 with serine, and substitution of the amino acid at position 35 with isoleucine;

(G-2) a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 11 having a substitution, deletion, or addition of 1 to 21 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 14, 18, 19, 26, 30, 33, and 34, and having a guanine-binding property;

(G-3) a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 40% to the sequence of SEQ ID NO: 11, provided that the amino acids at positions 1, 2, 3, 4, 6, 7, 9, 14, 18, 19, 26, 30, 33, and 34 are identical, and having a guanine-binding property, ;

(U-1) a polypeptide being a PPR motif consisting of the sequence of SEQ ID NO: 12, or a PPR motif consisting of the sequence of SEQ ID NO: 12 having a substitution selected from the group consisting of substitution of the amino acid at position 10 with phenylalanine, substitution of the amino acid at position 13 with serine, substitution of the amino acid at position 15 with lysine, substitution of the amino acid at position 17 with glutamic acid, substitution of the amino acid at position 20 with leucine, substitution of the amino acid at position 21 with lysine, substitution of the amino acid at position 23 with phenylalanine, substitution of the amino acid at position 24 with aspartic acid, substitution of the amino acid at position 27 with lysine, substitution of the amino acid at position 28 with lysine, substitution of the amino acid at position 29 with arginine, and substitution of the amino acid at position 31 with leucine;

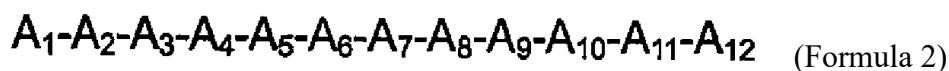
(U-2) a polypeptide being a PPR motif represented by formula 1 below consisting of the sequence of SEQ ID NO: 12 having a substitution, deletion, or addition of 1 to 22 amino acids other than the amino acids at positions 1, 2, 3, 4, 6, 11, 12, 14, 19, 26, 30, 33, and 34, and having a uracil-binding property; and

(U-3) a polypeptide being a PPR motif represented by formula 1 below having a sequence identity of at least 37% to the sequence of SEQ ID NO: 12, provided that the amino acids at positions 1, 2, 3, 4, 6, 11, 12, 14, 19, 26, 30, 33, and 34 are identical, and having a uracil-binding property.

(Helix A)-X-(Helix B)-L (Formula 1)

In the formula:

Helix A is a moiety of 12-amino acid length capable of forming an α -helix structure, and is represented by the formula 2;



wherein, in the formula 2, A₁ to A₁₂ independently represent an amino acid;

X does not exist, or is a moiety of 1- to 9-amino acid length;

Helix B is a moiety of 11- to 13-amino acid length capable of forming an α -helix structure; and

L is a moiety of 2- to 7-amino acid length represented by the formula 3;



wherein, in the formula 3, the amino acids are numbered "i" (-1), "ii" (-2), and so on from the C-terminus side,

provided that L_{iii} to L_{vii} may not exist.

[Claim 5]

Use of the polypeptide defined in any one of claims 1 to 3 or the polypeptide defined in claim 4 for preparation of a PPR protein, which is for enhancing binding performance of the PPR protein to a target RNA.

[Claim 6]

A protein comprising a polypeptide consisting of n of PPR motifs and capable of binding to a target RNA consisting of a sequence of n bases in length, wherein:

the PPR motif for adenine in the base sequence is the polypeptide of (A-1), (A-2'), or (A-3') defined in any one of claims 1 to 3, or the polypeptide of (A-1), (A-2), or (A-3) defined in claim 4;

the PPR motif for cytosine in the base sequence is the polypeptide of (C-1), (C-2'), or (c-3') defined in any one of claims 1 to 3 or the polypeptide of (C-1), (C-2), or (C-3) defined in claim 4;

the PPR motif for guanine in the base sequence is the polypeptide of (G-1), (G-2'), or (G-3') defined in any one of claims 1 to 3 or the polypeptide of (G-1), (G-2), or (G-3) defined in claim 4; and

the PPR motif for uracil in the base sequence is the polypeptide of (U-1), (U-2'), or (U-3) defined in any one of claims 1 to 3 or the polypeptide of (U-1), (U-2), or (U-3) defined in claim 4.

[Claim 7]

The protein according to claim 6, wherein n is 15 or larger.

[Claim 8]

A method for controlling RNA splicing, or a method for detecting RNA, which uses the protein according to claim 6 or 7.

[Claim 9]

A fusion protein of at least one selected from the group consisting of a fluorescent protein, a nuclear localization signal peptide, and a tag protein, and the protein according to claim 6 or 7.

[Claim 10]

A nucleic acid encoding the protein according to any one of claims 1 to 3, or the protein according to claim 6 or 7.

[Claim 11]

A vector comprising the nucleic acid according to claim 10.

[Claim 12]

A cell (except for human individual) containing the vector according to claim 11.

[Claim 13]

A method for manipulating RNA, which uses the protein according to any one of claims 1 to 3, the protein according to claim 6 or 7, or the vector according to claim 11 (implementation in human individual is excluded).

[Claim 14]

A method for producing an organism, which comprises the manipulation method according to claim 13.