CLAIMS

1. A nucleic acid encoding a protein that specifically binds to the CUG repeat sequence,

said protein that specifically binds to the CUG repeat sequence comprises at least 6 pentatricopeptide repeat (PPR) motifs consisting of polypeptides of 30 - 38 amino acids long represented by Formula 1,

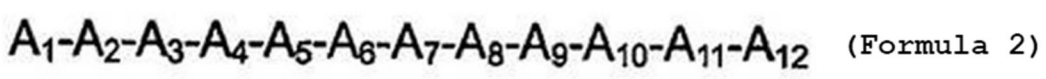
[Formula 1]



[wherein in Formula 1:

Helix A is a 12 amino acids long moiety that is capable of forming an α-helix structure, and is represented by Formula 2,

[Formula 2]



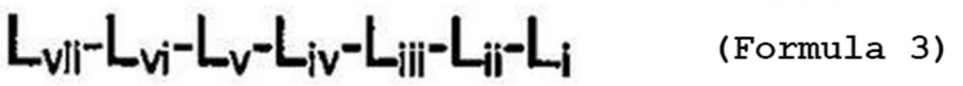
wherein in Formula 2, A_1 - A_{12} each independently represents an amino acid;

X is absent, or is a moiety consisting of a length of 1 - 9 amino acids;

Helix B is a moiety that is capable of forming an α-helix structure consisting of a length of 11 - 13 amino acids;

L is a moiety represented by Formula 3 that is 2 - 7 amino acids long;

[Formula 3]



and wherein in Formula 3, each amino acid is numbered from the C-terminal such as "i" (-1), "ii" (-2),

provided that L_{iii} - L_{vii} may be absent]

and the combination of three amino acids at A_1, A_4, and L_{ii}, or the combination of two amino acids at A_4 and L_{ii} in said each PPR motif is selected so that said each PPR motif binds to C, U, or G, and thereby configured so that said protein specifically binds to said CUG repeat sequence,

wherein the combination of three amino acids at A_1, A_4, and L_{ii} in said each PPR motif is:

when the base to be the target for the PPR motif is G (guanine), the combination of three amino acids at A_1, A_4, and L_{ii} is, in the order of (A_1, A_4, L_{ii}), (glutamic acid, glycine, aspartic acid), (valine, threonine, aspartic acid), (lysine, threonine, aspartic acid), or (leucine, threonine, aspartic acid);

when the base to be the target for the PPR motif is U (uracil), the combination of three amino acids at A_1, A_4, and L_{ii} is, in the order of (A_1, A_4, L_{ii}), (valine, asparagine, aspartic acid), (isoleucine, asparagine, asparagine), (isoleucine, asparagine, aspartic acid), (isoleucine, methionine, aspartic acid), (phenylalanine, proline, aspartic acid), or (tyrosine, proline, aspartic acid); or

when the base to be the target for the PPR motif is C (cytosine), the combination of three amino acids at A_1, A_4, and L_{ii} is, in the order of (A_1, A_4, L_{ii}), (valine, asparagine, asparagine), (isoleucine, asparagine, asparagine), (valine, asparagine, serine), or (isoleucine, methionine, aspartic acid);

and/or

wherein the combination of two amino acids at A_4 and L_{ii} in said each PPR motif is:

when the base to be the target for the PPR motif is G (guanine), the combination of two amino acids at A_4 and L_{ii} is, in the order of (A_4, L_{ii}), (threonine, aspartic acid), or (glycine, aspartic acid);

when the base to be the target for the PPR motif is U (uracil), the combination of two amino acids at A_4 and L_{ii} is, in the order of (A_4, L_{ii}), (asparagine, aspartic acid), (proline, aspartic acid), (methionine, aspartic acid), or (valine, threonine);

when the base to be the target for the PPR motif is C (cytosine), the combination of two amino acids at A_4 and L_{ii} is, in the order of (A_4, L_{ii}), (asparagine, asparagine), (asparagine, serine), or (leucine, aspartic acid).

2. The nucleic acid according to claim 1, wherein said protein that specifically binds to the CUG repeat sequence comprises 9 - 30 of said PPR motifs.

3. The nucleic acid according to claim 2, wherein said protein that specifically binds to the CUG repeat sequence comprises 12 - 24 of said PPR motifs.

4. The nucleic acid according to claim 1, characterized in that said nucleic acid encoding a protein that specifically binds to the CUG repeat sequence is integrated into an expression vector.

5. The nucleic acid according to claim 1, characterized in that said nucleic acid encoding a protein that specifically binds to the CUG repeat sequence is integrated into a viral vector.

6. The nucleic acid according to claim 5, characterized in that said viral vector is a viral vector with tropism towards muscle tissue.

7. The nucleic acid according to claim 6, characterized in that said viral vector is an adeno-associated virus (AAV) vector, an adenovirus vector, a retrovirus vector, a lentivirus vector, or a herpes simplex virus vector.

8. The nucleic acid according to claim 7, characterized in that said viral vector is an AAV vector.

9. The nucleic acid according to claim 8, characterized in that said AAV vector is an AAV1 vector, an AAV2 vector, an AAV6

vector, an AAV7 vector, an AAV8 vector, an AAV9 vector, an AAV10 vector, an AAV11 vector, or an AAV12 vector.

10. An expression vector comprising the nucleic acid encoding the protein that specifically binds to the CUG repeat sequence according to claim 1.

11. A cell comprising the expression vector according to claim 10.

12. A protein that specifically binds to the CUG repeat sequence that is produced from the cell according to claim 11.

13. A viral expression vector comprising the nucleic acid encoding the protein that specifically binds to the CUG repeat sequence as defined in claim 1.

14. A cell comprising the viral expression vector according to claim 13.

15. A viral vector that is produced from the cell according to claim 14.

16. The nucleic acid encoding the protein that specifically binds to the CUG repeat sequence according to any one of claims 1 to 9 for use in therapy.

17. The use of the nucleic acid encoding the protein that specifically binds to the CUG repeat sequence according to any one of claims 1 to 9 in the manufacture of a pharmaceutical composition for treating myotonic dystrophy type 1.