thermophiles is not yet available as SpCas9 is not active at or above 42°C (Mougiakos *et al. ACS Synth. Biol.* 6, 849–861 (2017)) and to date no thermophilic Cas9 has been characterized.

SUMMARY

The inventors have discovered and characterised ThermoCas9: an RNA-guided DNA-endonuclease from the CRISPR-Cas type-IIC system of the thermophilic bacterium *Geobacillus thermodenitrificans* T12. The inventors have surprisingly shown its *in vitro* activity over a broad temperature range, demonstrated the importance of the sgRNA-structure for thermostability and applied ThermoCas9 for *in vivo* genome editing across a wide temperature range.

In one aspect, the present disclosure refers to a nucleoprotein complex comprising a Cas protein, at least one targeting RNA molecule which recognises a target nucleic acid sequence in a double stranded target polynucleotide, and the target polynucleotide, wherein

the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence and a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; wherein the nucleoprotein complex is not in a human cell.

In another aspect, the present disclosure refers to a transformed non-human cell, having a double stranded target polynucleotide comprising a target nucleic acid sequence, wherein the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence, said cell comprising:

a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) protein having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

at least one targeting RNA molecule which recognizes the target nucleic acid sequence in the target nucleic acid strand, wherein the non-target strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; and

an expression vector comprising a nucleic acid encoding at least one of said Cas protein and said targeting RNA molecule.

In another aspect, the present disclosure refers to an *in vitro* method of binding, cleaving, marking or modifying a double stranded target polynucleotide, wherein the double stranded target polynucleotide comprises a target nucleic acid strand comprising a target nucleic acid sequence, and a non-target nucleic acid strand comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence, said method comprising:

a. designing at least one targeting RNA molecule, wherein the targeting RNA molecule recognizes the target sequence in the target strand, and the non-target strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3';

b. forming a ribonucleoprotein complex comprising the targeting RNA molecule and a Cas protein, wherein the isolated Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith; and

c. the ribonucleoprotein complex binding, cleaving, marking or modifying the target polynucleotide; and wherein said method is not used in human cells.

In yet another aspect, the present disclosure refers to the *in vitro* use of at least one targeting RNA molecule and a Cas protein to form a nucleoprotein complex for binding, cleaving, marking or modifying a double stranded target polynucleotide comprising a target nucleic acid sequence, wherein:

the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence;

the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

the at least one targeting RNA molecule recognizes the target sequence;

the non-target nucleic acid strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent to the 3' end of the protospacer nucleic acid sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; and wherein said use is not in a human cell.

Accordingly, the present disclosure provides an isolated clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) protein or polypeptide comprising;

a. the amino acid motif EKDGKYYC [SEQ ID NO: 2]; and/or
b. the amino acid motif $X_1X_2CTX_3X_4$ [SEQ ID NO: 3] wherein X_1 is independently selected from Isoleucine, Methionine or Proline, X_2 is independently selected from Valine, Serine, Asparagine or Isoleucine, X_3 is independently selected from Glutamate or Lysine and X_4 is one of Alanine, Glutamate or Arginine; and/or
c. the amino acid motif X_5LKX_6IE [SEQ ID NO: 4] wherein X_5 is independently selected from Methionine or Phenylalanine and X_6 is independently selected from Histidine or Asparagine; and/or
d. the amino acid motif X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine; and/or
e. the amino acid motif $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6] wherein X_9 is Alanine or Glutamate, X_{10} is Glutamine or Lysine, X_{11} is Arginine or Alanine, X_{12} is Asparagine or Alanine and X_{13} is Lysine or Serine.

For the avoidance of doubt proteins, polypeptides or nucleic acids coding for Cas proteins of the present disclosure may also be referred to as "GtCas9" or

CLAIMS

1. A nucleoprotein complex comprising a Cas protein, at least one targeting RNA molecule which recognises a target nucleic acid sequence in a double stranded target polynucleotide, and the target polynucleotide, wherein

the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence and a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; wherein the nucleoprotein complex is not in a human cell.

2. A transformed non-human cell, having a double stranded target polynucleotide comprising a target nucleic acid sequence, wherein the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence, said cell comprising:

a clustered regularly interspaced short palindromic repeat (CRISPR)-associated (Cas) protein having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

at least one targeting RNA molecule which recognizes the target nucleic acid sequence in the target nucleic acid strand, wherein the non-target strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; and

an expression vector comprising a nucleic acid encoding at least one of said Cas protein and said targeting RNA molecule.

3. A transformed cell as claimed in claim 2, wherein the Cas protein is expressed from an expression vector.

4. A nucleoprotein complex as claimed in claim 1, or a transformed cell as claimed in claim 2 or claim 3, wherein the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 95% identity therewith.

5. A nucleoprotein complex as claimed in claim 1 or claim 4, or transformed cell as claimed in any of claims 2 to 4, wherein the nucleoprotein complex occurs at a temperature between 20°C and 100°C, at a temperature between 30°C and 80°C, or at a temperature between 37°C and 78°C, preferably at a temperature above 55°C; more preferably at a temperature between 55°C and 80°C; even more preferably at a temperature between 55°C and 65°C or 60°C and 65°C.

6. A nucleoprotein complex or a transformed cell as claimed in any preceding claim, wherein the double stranded target polynucleotide comprising the target nucleic acid sequence is cleaved by the Cas protein; preferably wherein said cleavage is DNA cleavage.

7. A transformed cell as claimed in any of claims 2 to 5, wherein (i) binding, cleaving, marking or modifying of the double stranded target polynucleotide is at a temperature between 20°C and 100°C, at a temperature between 30°C and 80°C, or at a temperature between 37°C and 78°C; preferably at a temperature above 55°C; more preferably at a temperature between 55°C and 80°C; even more preferably at a temperature between 55°C and 65°C or 60°C and 65°C; and/or (ii) the target polynucleotide comprising the target sequence is double stranded DNA and said cleaving results in a double stranded break in the target polynucleotide.

8. A transformed cell or a nucleoprotein complex as claimed in any of claims to 5, wherein, the Cas protein lacks nuclease activity to cut the double stranded DNA yet there is gene silencing of the target polynucleotide.

9. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, wherein the PAM sequence comprises 5'-NNNNCNNA-3' [SEQ ID NO: 47]; preferably wherein the PAM sequence comprises 5'-NNNNCSAA-3' [SEQ ID NO: 48]; more preferably wherein the PAM sequence comprises 5'-NNNNCCAA-3' [SEQ ID NO: 50].

10. A nucleoprotein complex as claimed in claim 1 or claim 4, or a transformed cell as claimed in any of claims 2 to 4, wherein the nucleoprotein complex forms at a temperature between 20°C and 70°C; preferably at a temperature between 25°C and 65°C.

11. A transformed cell as claimed in any of claims 3 to 5, wherein binding, cleaving, marking or modifying of the double stranded target polynucleotide occurs at a temperature between 20°C and 70°C; preferably at a temperature between 25°C and 65°C.

12. A transformed cell or a nucleoprotein complex as claimed in any preceding claim wherein the cell is a prokaryotic cell or wherein the nucleoprotein complex is in a prokaryotic cell.

13. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, wherein the targeting RNA molecule comprises a crRNA and a tracrRNA.

14. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, wherein the length of the at least one targeting RNA molecule is in the range 35 – 200 nucleotide residues; and/or wherein the target nucleic acid sequence is from 15 to 32 nucleotide residues in length.

15. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, wherein the nuclease activity of the Cas protein is optionally inactive and the Cas protein further comprises at least one functional moiety; preferably a protein; optionally selected from a helicase, a nuclease, a helicase-nuclease, a DNA methylase, a histone methylase, an acetylase, a phosphatase, a kinase, a transcription (co-)activator, a transcription repressor, a DNA binding protein, a DNA structuring protein, a marker protein, a reporter protein, a fluorescent protein, a ligand binding protein, a signal peptide, a subcellular localisation sequence, an antibody epitope or an affinity purification tag, for example a green fluorescent protein (GFP).

16. An *in vitro* method of binding, cleaving, marking or modifying a double stranded target polynucleotide, wherein the double stranded target polynucleotide comprises a target nucleic acid strand comprising a target nucleic acid sequence,

and a non-target nucleic acid strand comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence, said method comprising:

a. designing at least one targeting RNA molecule, wherein the targeting RNA molecule recognizes the target sequence in the target strand, and the non-target strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3';

b. forming a ribonucleoprotein complex comprising the targeting RNA molecule and a Cas protein, wherein the isolated Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith; and

c. the ribonucleoprotein complex binding, cleaving, marking or modifying the target polynucleotide; and wherein said method is not used in human cells.

17. *In vitro* use of at least one targeting RNA molecule and a Cas protein to form a nucleoprotein complex for binding, cleaving, marking or modifying a double stranded target polynucleotide comprising a target nucleic acid sequence, wherein:

the double stranded target polynucleotide comprises a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence complementary to the target nucleic acid sequence;

the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 89% identity therewith;

the at least one targeting RNA molecule recognizes the target sequence;

the non-target nucleic acid strand further comprises a protospacer adjacent motif (PAM) sequence directly adjacent to the 3' end of the protospacer nucleic acid sequence, wherein the PAM sequence comprises 5'-NNNNCNN-3'; and wherein said use is not in a human cell.

18. An *in vitro* method as claimed in claim 16 or an *in vitro* use as claimed in claim 17, wherein the Cas protein has an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 95% identity therewith.

19. An *in vitro* method as claimed in claim 16 or claim 18, or an *in vitro* use as claimed in claim 17 or claim 18, wherein the nucleoprotein complex forms at a temperature between 20°C and 100°C, at a temperature between 30°C and 80°C, or at a temperature between 37°C and 78°C, preferably at a temperature above 55°C; more preferably at a temperature between 55°C and 80°C; even more preferably at a temperature between 55°C and 65°C or 60°C and 65°C.

20. An *in vitro* method as claimed in any of claims 16, 18 or 19, or an *in vitro* use as claimed in any of claims 17 to 19, wherein the PAM sequence comprises 5'-NNNNCNNA-3', 5'-CNNNCNN-3', 5'-NNNCCNN-3', 5'-NNCNCNN-3', 5'-NNNNCCN-3', and/or 5'-NCNNCNN-3'; preferably wherein the PAM sequence comprises 5'-CCCCCCNA-3' [SEQ ID NO: 10], more preferably wherein the PAM sequence comprises 5'-CCCCCCAA-3' [SEQ ID NO: 11].

21. An *in vitro* method as claimed in any of claims 16, 18, 19 or 20, or an *in vitro* use as claimed in any of claims 17 to 20, wherein the Cas protein further comprises at least one functional moiety; preferably a protein; optionally selected from a helicase, a nuclease, a helicase-nuclease, a DNA methylase, a histone methylase, an acetylase, a phosphatase, a kinase, a transcription (co-)activator, a transcription repressor, a DNA binding protein, a DNA structuring protein, a marker protein, a reporter protein, a fluorescent protein, a ligand binding protein, a signal peptide, a subcellular localisation sequence, an antibody epitope or an affinity purification tag, for example a green fluorescent protein (GFP).