example, microorganisms that are used in industrial fermentations or for *in vitro* laboratory processes conducted at elevated temperatures.

To date there is no experimental evidence for active Cas9 proteins in thermophiles. Based on a comparative genome screening by Chylinski *et al*. (2014; Nucleic Acids Research 42: 6091-61-05) on the presence of Cas9 in bacteria it was found that the Type II-C CRISPR-Cas system is only present in approximately 3.3% of all bacterial genomes. Among thermophilic bacteria, the Type II system is underrepresented based on statistical analysis (P=0.0019). In addition, no Type II system has been found in archaea however, this could possibly be due to the absence of the RNase III protein (involved in the Type II system) in archaea. Chylinski, *et al*., (2014; Nucleic Acids Research 42: 6091-6105) did describe the classification and evolution of type II CRISPR-Cas systems, in particular, two species are identified which exhibit these systems, however these species grow maximally at 55°C and do not exhibit strictly thermophilic growth with optimum growth temperature 60-80°C, with hyperthermophiles capable of growing optimally above 80°C.

Despite the rarity of the CRISPR-Cas system in bacterial genomes and in particular the fact that Cas9 has been found only in bacteria (not archaea) with optimal growth temperatures below 45°C, the inventors have surprisingly discovered several thermostable Cas9 variants which enable genome editing to be carried out at elevated temperatures. The inventors have also discovered optimised protospacer adjacent motif (PAM) sequences that work with the thermostable Cas9 variants to enable genome editing to be carried out over a wide range of temperatures, including at the elevated temperatures. These Cas9 nucleases, and RNA molecules that are designed with knowledge of the associated PAM sequences, provide novel tools for genetic engineering at elevated temperatures and are of particular value in the genetic manipulation of thermophilic organisms; particularly microorganisms.

SUMMARY

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 77% 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'.

In another aspect, the present disclosure refers to a transformed 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 77% 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; wherein the Cas protein and targeting RNA molecule enable binding, cleaving, marking or modifying of the target polynucleotide in the cell.

In yet another aspect, the present disclosure refers to a 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 77% identity therewith; and

c. the ribonucleoprotein complex binding, cleaving, marking or modifying the target polynucleotide.

In yet another aspect, the present disclosure refers to the use of at least one targeting RNA molecule and a Cas protein 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 77% 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'.

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.

A polypeptide in the context of this disclosure may be viewed as a fragment of the full length Cas protein. Such fragments may be inactive and used in ways and for purposes not associated directly with binding, editing and/or cutting of genetic material, for example for standards in assays or raising antibodies or the like.

In preferred examples however, the Cas protein or polypeptide is functional and capable of cleavage, binding, marking or modifying at a temperature in the range 20°C and 100°C, inclusive, when associated with at least one targeting RNA molecule, and a polynucleotide comprising a target nucleic acid sequence recognised by the targeting RNA molecule. Preferably the Cas protein or polypeptide is functional and capable of said cleavage, binding, marking or modifying at a temperature in the range 50°C and 70°C, for example 55°C or 60°C.

In particular examples, the disclosure may provide a Cas protein or polypeptide comprising the amino acid motif EKDGKYYC [SEQ ID NO: 2]. In other examples, the Cas proteins or polypeptides may further comprise 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.

In other examples the Cas proteins or polypeptides defined herein may additionally further comprise 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.

In other examples, the Cas proteins or polypeptides defined herein may additionally further comprise 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.

In other examples, the Cas proteins or polypeptides defined herein may additionally further comprise 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.

In accordance with the present disclosure, it may be appreciated that a Cas protein or polypeptide of the disclosure may comprise any of the motifs of SEQ ID NOs 2 to 6, either alone or in combination. The following summarises each of the combinations of motifs which may characterize Cas proteins or polypeptides of the disclosure:

EKDGKYYC [SEQ ID NO: 2].

EKDGKYYC [SEQ ID NO: 2]; and $X_1X_2CTX_3X_4$ [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.

EKDGKYYC [SEQ ID NO: 2]; and $X_1X_2CTX_3X_4$ [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

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 X_7VYSX_8K [SEQ ID NO: 5] wherein X_7 is Glutamate or Isoleucine and X_8 is one of Tryptophan, Serine or Lysine.

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 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [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.

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 $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [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.

In another example, the present disclosure provides an isolated Cas protein or polypeptide fragment thereof having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 77% identity therewith. Preferably the Cas protein or polypeptide is capable of binding, cleavage, marking or modifying at a temperature in the range 20°C and 100°C inclusive. Preferably the Cas protein or polypeptide is capable of said cleavage, binding, marking or modifying at a temperature in the range between 50°C and 70°C, for example 55°C or 60°C. Preferably the Cas protein or polypeptide is capable of said cleavage, binding, marking or modifying at a temperature in the range between 30°C and 80°C, 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.

The present disclosure also provides uses of a targeting RNA molecule and a Cas protein or polypeptide provided herein, for binding, cleaving, marking or modifying a target polynucleotide comprising a target nucleic acid sequence. The targeting RNA molecule recognizes the target nucleic acid sequence on a target nucleic acid strand of the polynucleotide.

The target polynucleotide that comprises the target nucleic acid sequence may be double stranded and so comprise a target nucleic acid strand, comprising said target nucleic acid sequence, and a non-target nucleic acid strand, comprising a protospacer nucleic acid sequence. The protospacer nucleic acid sequence is substantially complementary to the target nucleic acid sequence and pairs with it in the double stranded target polynucleotide. The non-target nucleic acid strand may further comprise a protospacer adjacent motif (PAM) sequence directly adjacent the 3' end of the protospacer sequence. The PAM sequence may be at least 6, 7, or 8 nucleic acids in length. Preferably, the PAM sequence has a cytosine in the fifth position. Preferably the PAM sequence comprises the sequence 5'-NNNNC-3', so that from the 5'-end the PAM sequence begins 5'-NNNNC-3'. Additionally or alternatively, the PAM sequence may have an adenine in the eighth position, so that the PAM sequence comprises the sequence 5'-NNNNNNNA-3', and from the 5'-end the PAM sequence begins 5'-NNNNNNNA-3'. Additionally or alternatively, the PAM sequence may have a cytosine in one or more of the first, second, third, fourth, and sixth positions, such that from the 5'-end the PAM sequence begins 5'-CNNNN-3', 5'-NCNNN-3', 5'-NNCNN-3', 5'-NNNCN-3', and/or 5'-NNNNNC-3'. Preferably the PAM sequence comprises, so that from the 5'-end the PAM sequence begins, 5'-CCCCCCNA-3' [SEQ ID NO: 10], and further preferably the PAM sequence comprises, so that from the 5'-end the PAM sequence begins, 5'-CCCCCCAA-3' [SEQ ID NO: 11]. Other preferred PAM sequences include 5'-ATCCCCAA-3' [SEQ ID NO: 21] and 5'-ACGGCCAA-3' [SEQ ID NO: 22].

Preferably, the Cas protein or polypeptide is capable of the binding, cleaving, marking or modifying at a temperature in the range 40°C to 80°C inclusive, preferably in the range 45°C to 80°C inclusive, and further preferably in the range 50°C to 80°C inclusive. For example, the binding, cleaving, marking or modifying occurs at a temperature of 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, 70°C, 71°C, 72°C, 73°C, 74°C, 75°C, 76°C, 77°C, 78°C, 79°C or 80°C. More preferably the Cas protein or polypeptide is capable of the binding, cleaving, marking or modifying at a temperature in the range 55 to 65°C. In preferred examples, a Cas protein or polypeptide fragment of the disclosure may comprises an amino acid

sequence of at least 75% identity; preferably at least 85%; more preferably at least 90%; even more preferably at least 95% identity to SEQ ID NO: 1.

The Cas protein or polypeptide may be used in combination with a targeting RNA molecule that recognizes a target nucleic acid sequence on the target nucleic acid strand, where the non-target nucleic acid sequence has a PAM sequence directly adjacent the 3' end of the protospacer sequence on the non-target strand, as disclosed herein. Thus, the PAM sequence may comprise the sequence 5'-NNNNC-3', and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Preferably from the 5'-end the PAM sequence begins 5'-NNNNC-3' and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Preferably from the 5'-end the PAM sequence begins 5'-NNNNNNNA-3' and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, the range 550°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive. Further preferably the 5'-end of the PAM sequence begins 5'-NNNNCNNA-3' and the Cas protein may bind, cleave, mark or modify the target strand at a temperature in the range 20°C and 100°C inclusive, preferably in the range 30°C and 90°C inclusive, in the range 37°C and 78°C inclusive, in the range 40°C and 80°C inclusive, in the range 50°C and 70°C inclusive, or in the range 55°C and 65°C, inclusive.

More particularly, a Cas protein or polypeptide of the disclosure may comprise an amino acid sequence with a percentage identity with SEQ ID NO:1 as follows: at least 60%, at least 61%, at least 62%, at least 63%, at least 64%, at least 65%, at least 66%, at least 67%, at least 68%, at least 69%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at

least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, at least 99%, at least 99.5% or at least 99.8%. The percentage identity may be at least 89%. The percentage identity may be at least 90%. Preferably the percentage identity will be at least 95%, for example 98%.

The percentage amino acid sequence identity with SEQ ID NO: 1 is determinable as a function of the number of identical positions shared by the sequences in a selected comparison window, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

A Cas protein or polypeptide fragment of the disclosure may be characterised in terms of both the reference sequence SEQ ID NO: 1 and any aforementioned percentage variant thereof as defined by percentage sequence identity, alone or in combination with any of the aforementioned amino acid motifs (i.e. SEQ ID NOS 2 and/or 3 and/or 4 and/or 5 and/or 6) as essential features.

The disclosure provides a use of a targeting RNA molecule as provided herein and a Cas protein or polypeptide of the disclosure for binding, cleaving, marking or modifying a target nucleic acid strand comprising a target nucleic acid sequence. Preferably said binding, cleaving, marking or modifying occurs at a temperature disclosed herein, for example at a temperature of between 20 and 100°C. The disclosure also provides a method of binding, cleaving, marking or modifying a target nucleic acid sequence in a target nucleic acid strand comprising designing a targeting RNA molecule as provided herein and forming a ribonucleoprotein complex comprising the targeting RNA molecule and a Cas protein or polypeptide of the disclosure. Preferably the ribonucleoprotein complex binding, cleaving, marking or modifying the target nucleic acid sequence at a temperature disclosed herein, for example at a temperature of between 37 and 100°C.

The uses and methods of the disclosure may be carried out, and the nucleoproteins of the disclosure formed and used, *in vivo*, for example in bacterial cells. Alternatively the uses and methods of the disclosure may be carried out, and the nucleoproteins of the disclosure formed and used, *in vitro*. The Cas protein of the disclosure may be provided

in isolated form, for example when used *in vitro* or when added to cells by transfection, the Cas protein may be heterologously expressed, for example following transient or stable transformation of the cell by nucleic acid encoding the Cas protein, the targeting RNA molecule may be transcribed from an expression vector following transient or stable transformation of the cell by nucleic acid encoding the RNA molecule, and/or the RNA molecule may be provided in isolated form, for example when used *in vitro* or when added to cells by transfection. In preferred examples, the Cas protein or polypeptide is expressed from the genome of a host cell, following stable intergration of a nucleic acid encoding the Cas protein or polypeptide in the genome of the host cell. Thus the Cas protein and/or RNA molecule may be added to the *in vivo* or *in vitro* environment using any artificial or contrived method for adding a protein or nucleic acid molecule to a cell in which it is not otherwise present.

The polynucleotide comprising the target nucleic acid sequence may be cleaved by the Cas protein, and optionally the cleavage may be DNA cleavage. The target nucleic acid strand comprising the target sequence may be double stranded DNA and the method or use may result in a double stranded break in the polynucleotide comprising the target nucleic acid sequence. The polynucleotide comprising the target nucleic acid sequence may be double stranded DNA, the Cas protein may lack the ability to cut the double stranded DNA and the use or method may result in gene silencing of the polynucleotide.

The Cas protein or polypeptide may be provided for the methods, uses and nucleoproteins of the disclosure at a concentration of 250 nM or less, for example at a concentration of 200 nM or less, 150 nM or less, 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. Alternatively, the Cas protein or polypeptide may be provided at a concentration of at least 0.5 nM, at least 1nM, at least 5nM, at least 10nM, at least 25nM, at least 50 nM, at least 100 nM, at least 150 nM, at least 200 nM, or at least 250 nM. The PAM sequence of the disclosure may have an adenine in the eighth position, so that the PAM sequence comprises the sequence 5'-NNNNNNNA-3', and the concentration of Cas protein or polypeptide may be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. The PAM sequence may comprise the sequence 5'-NNNNCNNA-3', and the concentration of Cas protein or polypeptide may

be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less. The PAM sequence may comprise the sequence 5'-CCCCCCNA-3' [SEQ ID NO: 10], and the concentration of Cas protein or polypeptide may be 100 nM or less, 50 nM or less, 25nM or less, 10 nM or less, 5nM or less, 1nM or less or 0.5 nM or less.

Also, the disclosure provides nucleic acids encoding any of the aforementioned proteins or polypeptides of the disclosure. The nucleic acids may be isolated or in the form of expression constructs.

In all aforementioned examples of the present disclosure, amino acid residues may be substituted conservatively or non-conservatively. Conservative amino acid substitutions refer to those where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not alter the functional properties of the resulting polypeptide.

Similarly it will be appreciated by a person of average skill in the art that nucleic acid sequences may be substituted conservatively or non-conservatively without affecting the function of the polypeptide. Conservatively modified nucleic acids are those substituted for nucleic acids which encode identical or functionally identical variants of the amino acid sequences. It will be appreciated by the skilled reader that each codon in a nucleic acid (except AUG and UGG; typically the only codons for methionine or tryptophan, respectively) can be modified to yield a functionally identical molecule. Accordingly, each silent variation (i.e. synonymous codon) of a polynucleotide or polypeptide, which encodes a polypeptide of the present disclosure, is implicit in each described polypeptide sequence.

The disclosure provides a transformed cell, having a target nucleic acid sequence in a double stranded target polynucleotide, said cell comprising a Cas protein or polypeptide as provided herein and at least one targeting RNA molecule as provided herein, and an expression vector comprising a nucleic acid encoding at least one of said Cas protein and said targeting RNA molecule. The Cas protein and targeting RNA molecule may enable or permit binding, cleaving, marking or modifying of the target sequence to occur in the transformed cell at a raised temperature, or at a range of

temperatures, for example between 37 and 100°C, as disclosed herein. The disclosure further provides a method of binding, cleaving, marking or modifying a target nucleic acid in a cell comprising either 1) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure and a nucleotide sequence encoding a targeting RNA molecule of the disclosure; or 2) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure and a further expression vector comprising a nucleotide sequence encoding a targeting RNA molecule of the disclosure; or 3) transforming, transfecting or transducing the cell with an expression vector comprising a nucleotide sequence encoding a Cas protein or polypeptide of the disclosure, and delivering a targeting RNA molecule as provided herein to, or into the cell. The Cas protein or polypeptide may be expressed from the genome of the transformed cell, for example following stable integration into the genome of a nucleotide sequence encoding the Cas protein or polypeptide.

The disclosure also provides kits comprising one or more of the reagents for carrying out the uses and methods of the disclosure, or for generating the transformed cells or nucleoprotein complex of the disclosure, said kits including: a Cas protein or polypeptide of the disclosure or an expression vector comprising a nucleic acid sequence encoding a Cas protein or polypeptide of the disclosure; and/or a targeting RNA molecule of the disclosure or an expression vector comprising a nucleic acid sequence encoding a targeting RNA molecule of the disclosure. The kits may further include instructions for carrying out the disclosure, for example instructions for how to design a targeting RNA molecule in accordance with the disclosure.

RNA Guides and Target Sequences

Cas proteins of the invention allow for sequence-specific binding, cleavage, tagging, marking or modification of target nucleic acids at elevated temperatures. Target nucleic acids may be DNA (single-stranded or double-stranded), RNA or synthetic nucleic acids. A particularly useful application of the present invention is the

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 77% 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'.

2. A transformed 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 77% 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; wherein the Cas protein and targeting RNA molecule enable binding, cleaving, marking or modifying of the target polynucleotide in the cell.

3. A transformed cell as claimed in claim 1, 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 binding, cleaving, marking or modifying, or the nucleoprotein complex forms at a temperature between 20°C and 100°C, at a temperature between 30°C and 80°C, 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.

5. A nucleoprotein complex or 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.

6. A nucleoprotein complex or transformed cell as claimed in any preceding claim, wherein the target polynucleotide comprising the target sequence is double stranded DNA and said binding, cleaving, marking or modifying results in a double stranded break in the target polynucleotide.

7. A transformed cell or nucleoprotein complex as claimed in any of claims 1 to 4, wherein the Cas protein lacks nuclease activity to cut the double stranded DNA, yet there is gene silencing of the target polynucleotide.

8. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, 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].

9. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, wherein the nucleoprotein complex is in a prokaryotic cell; preferably a thermophilic bacterium.

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

11. 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.

12. A transformed cell or a nucleoprotein complex as claimed in any preceding claim, 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).

13. A 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 77% identity therewith; and

c. the ribonucleoprotein complex binding, cleaving, marking or modifying the target polynucleotide.

14. Use of at least one targeting RNA molecule and a Cas protein 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 77% 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'.

15. A method as claimed in claim 14 or a use as claimed in claim 15, further comprising any of the features as set forth in any of claims 4 to 12.