In the fields of genetic engineering and genomic research there is an ongoing need for improved agents for sequence/site specific nucleic acid detection and/or cleavage.

The inventors have made a surprising discovery in that certain bacteria expressing Cas3, which has helicase-nuclease activity, express Cas3 as a fusion with Cse1. The inventors have also unexpectedly been able to produce artificial fusions of Cse1 with other nuclease enzymes.

The inventors have also discovered that Cas3-independent target DNA recognition by Cascade marks DNA for cleavage by Cas3, and that Cascade DNA binding is governed by topological requirements of the target DNA.

The inventors have further found that Cascade is unable to bind relaxed target plasmids, but surprisingly Cascade displays high affinity for targets which have a negatively supercoiled (nSC) topology.

Accordingly ~~in a first aspect~~ the present ~~invention~~ disclosure provides a clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defence (Cascade), the Cascade protein complex, or portion thereof, comprising at least CRISPR-associated protein subunits:

- Cas7 (or COG 1857) having an amino acid sequence of SEQ ID NO:3 or a sequence of at least 18% identity therewith,
- Cas5 (or COG1688) having an amino acid sequence of SEQ ID NO:4 or a sequence of at least 17% identity therewith, and
- Cas6 (or COG 1583) having an amino acid sequence of SEQ ID NO:5 or a sequence of at least 16% identity therewith;

and wherein at least one of the subunits includes an additional amino acid sequence providing nucleic acid or chromatin modifying, visualising, transcription activating or transcription repressing activity.

According to one aspect, there is provided an artificial fusion of (i) at least one clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein subunit, and (ii) a nuclease, wherein the CRISPR-associated protein subunit is selected from the group consisting of Cas6, Cas5, Cse1, Cse2 and Cas7.

According to another aspect, there is provided a nucleic acid molecule encoding a fusion as defined herein.

According to another aspect, there is provided an expression vector comprising a nucleic acid molecule as defined herein.

According to another aspect, there is provided a Cascade protein complex comprising a fusion as defined herein.

According to another aspect, there is provided a ribonucleoprotein complex comprising a fusion as defined herein and a CRISPR RNA (crRNA) molecule.

According to another aspect, there is provided a eukaryotic cell comprising a ribonucleoprotein complex as defined herein, wherein the cell is other than a human embryonic stem cell.

According to another aspect, there is provided a method of modifying a target nucleic acid *in vitro* comprising contacting the target nucleic acid with a ribonucleoprotein complex as defined herein.

CLAIMS

1. A clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defence (Cascade), the Cascade protein complex, or portion thereof, comprising at least CRISPR-associated protein subunits:
   - Cas7 having an amino acid sequence of SEQ ID NO:3 or a sequence of at least 18% identity therewith,
   - Cas5 having an amino acid sequence of SEQ ID NO:4 or a sequence of at least 17% identity therewith, and
   - Cas6 having an amino acid sequence of SEQ ID NO:5 or a sequence of at least 16% identity therewith;

   and wherein at least one of the subunits includes an additional amino acid sequence providing nucleic acid or chromatin modifying, visualising, transcription activating or transcription repressing activity.

2. A protein complex as claimed in claim 1, wherein the Cas6 subunit is a Cas6e subunit having an amino acid sequence of SEQ ID NO: 17 or a sequence of at least 16% identity therewith.

3. A protein complex as claimed in claim 1 or claim 2, further comprising a Cse2 subunit having an amino acid sequence of SEQ ID NO:2 or a sequence of at least 20% identity therewith; optionally wherein it is the Cse2 subunit which includes the additional amino acid sequence.

4. A protein complex as claimed in any of claims 1 to 3, further comprising a Cse1 subunit having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 9% identity therewith; optionally wherein it is the Cse1 subunit which includes the additional amino acid sequence.

5. A protein complex as claimed in claim 4 which is a Type I CRISPR-Cas system protein complex; preferably a subtype I-E CRISPR-Cas protein complex.

6. A protein complex as claimed in any of claims 1 to 5, wherein the additional amino acid sequence is translationally fused or covalently linked to the at least one subunit; preferably the additional amino acid sequence is fused or linked to at least the N

terminus and/or the C terminus of at least one of a Cse1, Cse2, Cas7, Cas5, Cas6 or Cas6e subunit.

7. A protein complex as claimed in claim 6, wherein the additional amino acid sequence is fused or linked to the N terminus or the C terminus of a Cse1, Cse2 or Cas5 subunit; preferably the N terminus of a Cse1 subunit, the N terminus of a Cse2 subunit, or the N terminus of a Cas7 subunit.

8. A protein complex as claimed in claim 7, wherein the additional amino acid sequence is a protein; optionally selected from a helicase, a nuclease, a nuclease-helicase (e.g. Cas3), a DNA methyltransferase (e.g. Dam), or DNA demethylase, a histone methyltransferase, a histone demethylase, an acetylase, a deacetylase, a phosphatase, a kinase, a transcription (co-)activator, an RNA polymerase subunit, a transcription repressor, a DNA binding protein, a DNA structuring protein, a marker protein, a reporter protein, a fluorescent protein, a ligand binding protein (e.g. mCherry or a heavy metal binding protein), a signal peptide (e.g. Tat-signal sequence), a subcellular localisation sequence (e.g. nuclear localisation sequence), or an antibody epitope.

9. A protein complex as claimed in claim 8, wherein the nuclease is selected from a type II restriction endonuclease; preferably FokI; more preferably a modified FokI, e.g. KKR Sharkey or ELD Sharkey.

10. A protein complex as claimed in any preceding claim further comprising an RNA molecule comprising a ribonucleotide sequence of at least 50% identity to a target nucleic acid sequence, and wherein the protein complex and the RNA molecule form a ribonucleoprotein complex.

11. A ribonucleoprotein complex as claimed in claim 10, wherein a portion of the RNA molecule has the at least 50% identity to the target sequence.

12. A ribonucleoprotein complex as claimed in claim 11, wherein the portion of the RNA molecule is at least substantially complementary along its length to the target sequence.

13. A ribonucleoprotein complex as claimed in any of claims 10 to 12, wherein the length of the RNA molecule is in the range 35 – 75 residues.

14. A ribonucleoprotein complex as claimed in any of claims 10 to 13, wherein the portion of the RNA molecule used to target a desired nucleic acid sequence is 32 or 33 residues long.

15. A ribonucleoprotein complex as claimed in any of claims 10 to 14, wherein the RNA molecule comprises 8 residues which are 5' to the RNA sequence which has at least substantial complementarity to the target sequence.

16. A ribonucleoprotein complex as claimed in any of claims 10 to 15, wherein the RNA molecule has a hairpin and tetranucleotide loop forming sequence 3' to the RNA sequence which has at least substantial complementarity to the target sequence.

17. An isolated nucleic acid molecule encoding at least one clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein subunit selected from:

a. a Cse1 subunit having an amino acid sequence of SEQ ID NO: 1 or a sequence of at least 9% identity therewith;
b. a Cse2 subunit having an amino acid sequence of SEQ ID NO:2 or a sequence of at least 20% identity therewith;
c. a Cas7 subunit having an amino acid sequence of SEQ ID NO:3 or a sequence of at least 18% identity therewith;
d. a Cas5 subunit having an amino acid sequence of SEQ ID NO:4 or a sequence of at least 17% identity therewith;
e. a Cas6 subunit having an amino acid sequence of SEQ ID NO:5 or a sequence of at least 16% identity therewith; and

wherein at least a, b, c, d or e includes an additional amino acid sequence providing nucleic acid or chromatin modifying, visualising, transcription activating or transcription repressing activity.

18. An isolated nucleic acid molecule as claimed in claim 17, wherein the additional amino acid sequence providing nucleic acid or chromatin modifying, visualising, transcription activating or transcription repressing activity is fused to the CRISPR-associated protein subunit.

19. An isolated nucleic acid molecule as claimed in claim 18, wherein the additional amino acid sequence is selected from a helicase, a nuclease, a nuclease-helicase (e.g. Cas3), a DNA methyltransferase (e.g. Dam), a DNA demethylase, a histone methyltransferase, a histone demethylase, an acetylase, a deacetylase, a phosphatase, a kinase, a transcription (co-)activator, an RNA polymerase subunit, a transcription repressor, a DNA binding protein, a DNA structuring protein, a marker protein, a reporter protein, a fluorescent protein, a ligand binding protein (e.g. mCherry or a heavy metal binding protein), a signal peptide (e.g. Tat-signal sequence), a subcellular localisation sequence (e.g. nuclear localisation sequence), or an antibody epitope.

20. An expression vector comprising a nucleic acid molecule of any of claims 17 to 19.

21. An expression vector as claimed in claim 20 further comprising a nucleotide sequence encoding an RNA molecule as defined in any of claims 10 to 16.

22. A method of modifying, visualising, activating transcription, or repressing transcription of a target nucleic acid comprising contacting the nucleic acid with:
    a. a ribonucleoprotein complex of any of claims 10 to 16; or
    b. a protein complex of any of claims 1 to 9, and an RNA molecule as defined in any of claims 10 to 16.

23. A method of modifying, visualising, activating transcription, or repressing transcription of a target nucleic acid in a cell, comprising transfecting, transforming or transducing the cell with an expression vector of claim 22, and a further expression vector comprising a nucleotide sequence encoding an RNA molecule as defined in any of claims 10 to 18.

24. A method of modifying, visualising, activating transcription, or repressing transcription of a target nucleic acid in a cell, comprising transfecting, transforming or

transducing the cell with an expression vector of claim 22, and then administering an RNA molecule as defined in any of claims 10 to 18 to or into the cell.

25. A method of modifying, visualising, activating transcription, or repressing transcription of a target nucleic acid in a cell, comprising transfecting, transforming or transducing the cell with an expression vector of claim 21.

26. A method of modifying or visualising a target nucleic acid as claimed in any of claims 22 to 25, wherein the additional amino acid sequence having nucleic acid or chromatin modifying or visualising activity, is a marker and the marker associates with the target nucleic acid or chromatin; preferably wherein the marker is a protein; optionally a fluorescent protein, e.g. green fluorescent protein (GFP) or yellow fluorescent protein (YFP).

27. A method as claimed in any of claims 22 to 26, wherein the target nucleic acid is DNA; preferably dsDNA.

28. A method as claimed in any of claims 22 to 26, wherein the target nucleic acid is RNA; preferably mRNA.

29. A method of modifying a target nucleic acid as claimed in any of claims 22 to 26, wherein the nucleic acid is dsDNA, the additional amino acid sequence having nucleic acid or chromatin modifying activity is a nuclease or a nuclease-helicase, and the modification is a single stranded or a double stranded break at a desired locus.

30. A method of non-homologous end joining of a dsDNA molecule in a cell at a desired locus to remove at least a part of a nucleotide sequence from the dsDNA molecule; optionally to knockout the function of a gene or genes, wherein the method comprises making double stranded breaks using a method of modifying a target nucleic acid as claimed in claim 29.

31. A method of homologous recombination of a nucleic acid into a dsDNA molecule in a cell at a desired locus in order to modify an existing nucleotide sequence or insert a desired nucleotide sequence, wherein the method comprises making a single stranded

or double stranded break at the desired locus using a method of modifying a target nucleic acid as claimed in claim 29.

32. A method of modifying, activating or repressing gene expression in an organism comprising modifying a target nucleic acid sequence as claimed in a method of any of claims 22 to 25, wherein the nucleic acid is dsDNA and the additional amino acid sequence having nucleic acid or chromatin modifying, transcription activating or repressing activity is selected from a DNA modifying enzyme (e.g. a demethylase or deacetylase), a transcription activator or a transcription repressor.

33. A method of modifying, activating or repressing gene expression in an organism comprising modifying a target nucleic acid sequence as claimed in a method of any of claims 22 to 25, wherein the nucleic acid is an mRNA and the additional amino acid sequence having nucleic acid or chromatin modifying or transcription activating or repressing activity is a ribonuclease; optionally selected from an endonuclease, a 3' exonuclease or a 5' exonuclease.

34. A method as claimed in any of claims 22 to 33, wherein the cell is a prokaryote.

35. A method as claimed in any of claims 22 to 33, wherein the cell is a eukaryotic cell, e.g. a plant cell, a yeast cell, a fungal cell, an insect cell, a mammalian cell or a human cell.

36. A method as claimed in claim 35, wherein the cell of the mammal or human is a stem cell other than a human embryonic stem cell; preferably an isolated stem cell.

37. A method as claimed in any of claims 33 to 35, wherein the cell is transfected *in vitro*.

38. A method as claimed in any of claims 22 to 36, wherein the target nucleic acid has a tertiary structure, optionally supercoiled, preferably wherein the target nucleic acid is negatively supercoiled.

39. A pharmaceutical composition comprising a Cascade protein complex of any of claims 1 to 9.

40. A pharmaceutical composition comprising a ribonucleoprotein complex of any of claims 10 to 16.

41. A pharmaceutical composition comprising an isolated nucleic acid of any of claims 17 to 19, or an expression vector of claim 20 or claim 21.

42. A kit comprising a Cascade protein complex of any of claims 1 to 9 and an RNA molecule as defined in any of claims 10 to 16.

43. A Cascade protein complex of any of claims 1 to 9 for use as a medicament.

44. A ribonucleoprotein complex of any of claims 10 to 16 for use as a medicament.

45. An isolated nucleic acid of any of claims 17 to 19 for use as a medicament.

46. An expression vector of claim 20 or claim 21 for use as a medicament.

CLAIMS:

1. An artificial fusion of (i) at least one clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein subunit, and (ii) a nuclease, wherein the CRISPR-associated protein subunit is selected from the group consisting of Cas6, Cas5, Cse1, Cse2 and Cas7.

2. A fusion according to claim 1 wherein the fusion further comprises a nuclear localization signal.

3. A fusion according to claim 1 or claim 2 wherein the nuclease comprises FokI or a mutant or an active portion thereof.

4. A fusion according to claim 3 wherein the nuclease is a modified FokI, which is KKR Sharkey or ELD Sharkey or both KKR Sharkey and ELD Sharkey together in combination.

5. A fusion according to any one of claims 1-4 wherein the CRISPR-associated protein subunit is Cse1.

6. A nucleic acid molecule encoding a fusion according to any one of claims 1-5.

7. An expression vector comprising a nucleic acid molecule according to claim 6.

8. A Cascade protein complex comprising a fusion according to any one of claims 1-5.

9. A ribonucleoprotein complex comprising a fusion according to any one of claims 1-5 and a CRISPR RNA (crRNA) molecule.

10. A eukaryotic cell comprising a ribonucleoprotein complex according to claim 9, wherein the cell is other than a human embryonic stem cell.

11. A method of modifying a target nucleic acid *in vitro* comprising contacting the target nucleic acid with a ribonucleoprotein complex according to claim 9.

12. A method according to claim 11 wherein the method of modifying is binding and/or cleaving the target nucleic acid.

13. A method according to claim 11 or claim 12, wherein the target nucleic acid is a double-stranded DNA (dsDNA).