(51) Int.Cl.: *C 12 N 9/22 (2006.01)*

(45) Oversættelsen bekendtgjort den: **2016-08-01**

(80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om meddelelse af patentet: **2016-04-06**

(86) Europæisk ansøgning nr.: **13165733.0**

(86) Europæisk indleveringsdag: **2008-10-31**

(87) Den europæiske ansøgnings publiceringsdag: **2013-11-06**

(30) Prioritet: **2007-10-31 US 1247 P**

(62) Stamansøgningsnr: **08845549.8**

(84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR**

(73) Patenthaver: **Precision Biosciences, Inc., 302 East Pettigrew Street , Dibrell Building, Suite A-100, Durham, NC 27701, USA**

(72) Opfinder: **SMITH, James, Jefferson, 1808 CARNATION DRIVE, Durham, NC 27703, USA**
JANTZ, Derek, 215 Hillview Drive, Durham, NC 27703, USA

(74) Fuldmægtig i Danmark: **Chas. Hude A/S, H.C. Andersens Boulevard 33, 1780 København V, Danmark**

(54) Benævnelse: **Rationelt designede enkeltkædede meganukleaser med ikke-palindromiske genkendelsessekvenser**

(56) Fremdragne publikationer:
WO-A1-03/078619
WO-A2-2007/047859
EPINAT J-C ET AL: "A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells", NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB LNKD-DOI:10.1093/NAR/GKG375, vol. 31, no. 11, 1 June 2003 (2003-06-01), pages 2952-2962, XP002248751, ISSN: 0305-1048
SMITH JULIANNE ET AL: "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences", NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 34, no. 22, 27 November 2006 (2006-11-27), pages E149-1, XP002457876, ISSN: 0305-1048
PRIETO JESUS ET AL: "The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage", NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB LNKD-DOI:10.1093/NAR/GKM183, vol. 35, no. 10, 22 April 2007 (2007-04-22), pages 3262-3271, XP009108728, ISSN: 0305-1048
ARNOULD S ET AL: "Engineering of Large Numbers of Highly Specific Homing Endonucleases that Induce Recombination on Novel DNA Targets", JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD-DOI:10.1016/J.JMB.2005.10.065, vol. 355, no. 3, 20 January 2006 (2006-01-20), pages 443-458, XP024950505, ISSN: 0022-2836 [retrieved on 2006-01-20]

DESCRIPTION

FIELD OF THE INVENTION

[0001] The invention relates to the field of molecular biology and recombinant nucleic acid technology. In particular, the invention relates to rationally-designed, non-naturally-occurring meganucleases in which a pair of enzyme subunits having specificity for different recognition sequence half-sites are joined into a single polypeptide to form a functional heterodimer with a non-palindromic recognition sequence. The invention also relates to methods of producing such meganucleases, and methods of producing recombinant nucleic acids and organisms using such meganucleases.

BACKGROUND OF THE INVENTION

[0002] Genome engineering requires the ability to insert, delete, substitute and otherwise manipulate specific genetic sequences within a genome, and has numerous therapeutic and biotechnological applications. The development of effective means for genome modification remains a major goal in gene therapy, agrotechnology, and synthetic biology (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73; Tzfira et al. (2005), Trends Biotechnol. 23: 567-9; McDaniel et al. (2005), Curr. Opin. Biotechnol. 16: 476-83). A common method for inserting or modifying a DNA sequence involves introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target and selecting or screening for a successful homologous recombination event. Recombination with the transgenic DNA occurs rarely but can be stimulated by a double-stranded break in the genomic DNA at the target site. Numerous methods have been employed to create DNA double-stranded breaks, including irradiation and chemical treatments. Although these methods efficiently stimulate recombination, the double-stranded breaks are randomly dispersed in the genome, which can be highly mutagenic and toxic. At present, the inability to target gene modifications to unique sites within a chromosomal background is a major impediment to successful genome engineering.

[0003] One approach to achieving this goal is stimulating homologous recombination at a double-stranded break in a target locus using a nuclease with specificity for a sequence that is sufficiently large to be present at only a single site within the genome (see, *e.g.,* Porteus et al. (2005), Nat. Biotechnol. 23: 967-73). The effectiveness of this strategy has been demonstrated in a variety of organisms using chimeric fusions between an engineered zinc finger DNA-binding domain and the non-specific nuclease domain of the FokI restriction enzyme (Porteus (2006), Mol. Ther. 13: 438-46; Wright et al. (2005), Plant J. 44: 693-705; Urnov et al. (2005), Nature 435: 646-51). Although these artificial zinc finger nucleases stimulate site-specific recombination, they retain residual non-specific cleavage activity resulting from under-regulation of the nuclease domain and frequently cleave at unintended sites (Smith et al. (2000), Nucleic Acids Res. 28: 3361-9). Such unintended cleavage can cause mutations and toxicity in the treated organism (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73).

[0004] A group of naturally-occurring nucleases which recognize 15-40 base-pair cleavage sites commonly found in the genomes of plants and fungi may provide a less toxic genome engineering alternative. Such "meganucleases" or "homing endonucleases" are frequently associated with parasitic DNA elements, such as group I self-splicing introns and inteins. They naturally promote homologous recombination or gene insertion at specific locations in the host genome by producing a double-stranded break in the chromosome, which recruits the cellular DNA-repair machinery (Stoddard (2006), Q. Rev. Biophys. 38: 49-95). Meganucleases are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cys box family and the HNH family. These families are characterized by structural motifs, which affect catalytic activity and recognition sequence. For instance, members of the LAGLIDADG family are characterized by having either one or two copies of the conserved LAGLIDADG motif (see Chevalier et al. (2001), Nucleic Acids Res. 29(18): 3757-3774). The LAGLIDADG meganucleases with a single copy of the LAGLIDADG motif ("mono-LAGLIDADG meganucleases") form homodimers, whereas members with two copies of the LAGLIDADG motif ("di-LAGLIDADG meganucleases") are found as monomers. Mono-LAGLIDADG meganucleases such as I-CreI, I-CeuI, and I-MsoI recognize and cleave DNA sites that are palindromic or pseudo-palindromic, while di-LAGLIDADG meganucleases such as I-SceI, I-AniI, and I-DmoI generally recognize DNA sites that are non-palindromic (Stoddard (2006), Q. Rev. Biophys. 38: 49-95).

[0005] Natural meganucleases from the LAGLIDADG family have been used to effectively promote site-specific genome modification in plants, yeast, *Drosophila,* mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the meganuclease recognition sequence (Monnat et al. (1999), Biochem. Biophys. Res. Common. 255: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Rouet et al. (1994), Mol. Cell. Biol. 14: 8096-106; Chilton et al. (2003), Plant Physiol. 133: 956-65; Puchta et al. (1996), Proc. Natl. Acad. Sci. USA 93: 5055-60; Rong et al. (2002), Genes Dev. 16: 1568-81; Gouble et al. (2006), J. Gene Med. 8(5):616-622).

[0006] Systematic implementation of nuclease-stimulated gene modification requires the use of genetically engineered enzymes with customized specificities to target DNA breaks to existing sites in a genome and, therefore, there has been great interest in adapting meganucleases to promote gene modifications at medically or biotechnologically relevant sites (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73; Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62).

[0007] I-CreI is a member of the LAGLIDADG family which recognizes and cuts a 22 base-pair recognition sequence in the chloroplast chromosome, and which presents an attractive target for meganuclease redesign. The wild-type enzyme is a homodimer in which each monomer makes direct contacts with 9 base pairs in the full-length recognition sequence. Genetic selection techniques have been used to

modify the wild-type I-CreI cleavage site preference (Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Chames et al. (2005), Nucleic Acids Res. 33: e178; Seligman et al. (2002), Nucleic Acids Res. 30: 3870-9, Arnould et al. (2006), J. Mol. Biol. 355: 443-58, Rosen et al. (2006), Nucleic Acids Res. 34: 4791-4800, Arnould et al. (2007). J. Mol. Biol. 371: 49-65, WO 2008/010009, WO 2007/093918, WO 2007/093836, WO 2006/097784, WO 2008/059317, WO 2008/059382, WO 2008/102198, WO 2007/060495, WO 2007/049156, WO 2006/097853, WO 2004/067736). More recently, a method of rationally-designing mono-LAGLIDADG meganucleases was described which is capable of comprehensively redesigning I -CreI and other such meganucleases to target widely-divergent DNA sites, including sites in mammalian, yeast, plant, bacterial, and viral genomes (WO 2007/047859).

[0008] A major limitation of using mono-LAGLIDADG meganucleases such as I-CreI for most genetic engineering applications is the fact that these enzymes naturally target palindromic DNA recognition sites. Such lengthy (10-40 bp) palindromic DNA sites are rare in nature and are unlikely to occur by chance in a DNA site of interest. In order to target a non-palindromic DNA site with a mono-LAGLIDADG meganuclease, one can produce a pair of monomers which recognize the two different half-sites and which heterodimerize to form a meganuclease that cleaves the desired non-palindromic site. Heterodimerization can be achieved either by co-expressing a pair of meganuclease monomers in a host cell or by mixing a pair of purified homodimeric meganucleases *in vitro* and allowing the subunits to re-associate into heterodimers (Smith et al. (2006), Nuc. Acids Res. 34:149-157; Chames et al. (2005), Nucleic Acids Res. 33:178-186; WO 2007/047859, WO 2006/097854, WO 2007/057781, WO 2007/049095, WO 2007/034262). Both approaches suffer from two primary limitations: (1) they require the expression of two meganuclease genes to produce the desired heterodimeric species (which complicates gene delivery and *in vivo* applications) and (2) the result is a mixture of approximately 25% the first homodimer, 50% the heterodimer, and 25% the second homodimer, whereas only the heterodimer is desired. This latter limitation can be overcome to a large extent by genetically engineering the dimerization interfaces of the two meganucleases to promote heterodimerization over homodimerization as described in WO 2007/047859, WO 2008/093249, WO 2008/093152, and Fajardo-Sanchez et al. (2008). Nucleic Acids Res. 36:2163-2173. Even so, two meganuclease genes must be expressed and homodimerization is not entirely prevented.

[0009] An alternative approach to the formation of meganucleases with non-palindromic recognition sites derived from one or more mono-LAGLIDADG meganucleases is the production of a single polypeptide which comprises a fusion of the LAGLIDADG subunits derived from two meganucleases. Two general methods can be applied to produce such a meganuclease.

[0010] In the first method, one of the two LAGLIDADG subunits of a di-LAGLIDADG meganuclease can be replaced by a LAGLIDADG subunit from a mono-LAGLIDADG meganuclease. This approach was demonstrated by replacing the C-terminal subunit of the di-LAGLIDADG I-DmoI meganuclease with an I-CreI subunit (Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62; Chevalier et al. (2002), Mol. Cell 10:895-905; WO 2003/078619). The result was a hybrid I-DmoI/I-CreI meganuclease which recognized and cleaved a hybrid DNA site.

[0011] In the second method, a pair of mono-LAGLIDADG subunits can be joined by a peptide linker to create a "single-chain heterodimer meganuclease." One attempt to produce such a single-chain derivative of I-CreI has been reported (Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62; WO 2003/078619). However, as discussed herein as well as in Fajardo-Sanchez et al. (2008), Nucleic Acids Res. 36:2163-2173 , there is now evidence suggesting that this method did not produce a single-chain heterodimer meganuclease in which the covalently joined I-CreI subunits functioned together to recognize and cleave a non-palindromic recognition site.

[0012] Therefore, there remains a need in the art for methods for the production of single-chain heterodimer meganucleases derived from mono-LAGLIDADG enzymes such as I-CreI to recognize and cut non-palindromic DNA sites.

SUMMARY OF THE INVENTION

[0013] The invention is defined by the appended claims.

[0014] The present invention is based, in part, upon the development of fusion proteins in which a peptide linker covalently joins two heterologous LAGLIDADG meganuclease subunits to form a "single-chain heterodimer meganuclease" or "single-chain meganuclease", in which at least the N-terminal subunit is derived from a mono-LAGLIDADG meganuclease, and in which the subunits function together to preferentially bind to and cleave a non-palindromic DNA recognition site which is a hybrid of the recognition half-sites of the two subunits. In particular, the invention can be used to genetically engineer single-chain meganucleases which recognize non-palindromic DNA sequences that naturally-occurring meganucleases do not recognize. The invention also provides methods that use such meganucleases to produce recombinant nucleic acids and organisms by utilizing the meganucleases to cause recombination of a desired genetic sequence at a limited number of loci within the genome of the organism for, *enter alia,* genetic engineering, gene therapy, treatment of pathogenic infections, and *in vitro* applications in diagnostics and research.

[0015] Thus, disclosed are recombinant single-chain meganucleases comprising a pair of covalently joined LAGLIDADG subunits derived from one or more mono-LAGLIDADG meganucleases which function together to recognize and cleave a non-palindromic recognition site. The mono-LAGLIDADG subunit can be derived from a wild-type meganuclease selected from I-CreI, I-MsoI and I-CeuI.

[0016] Disclosed are also recombinant single-chain meganucleases comprising a pair of mono-LAGLIDADG subunits in which the N-terminal subunit is derived from a wild-type meganuclease selected from I-CreI, I-MsoI and I-CeuI, and the C-terminal subunit is also derived from a

wild-type meganuclease selected from I-Crel, I-Msol and I-Ceul, but the N-terminal subunit is derived from a wild-type meganuclease of a different species than the C-terminal subunit.

[0017] Disclosed are also recombinant single-chain meganucleases comprising a pair of LAGLIDADG subunits in which the N-terminal subunit is derived from a wild-type meganuclease selected from I-Crel, I-Msol and I-Ceul, and the C-terminal subunit is derived from a single LAGLIDADG subunit from a wild-type di-LAGLIDADG meganuclease selected from I-Dmol, I-Scel and I-Anil.

[0018] Wild-type mono-LAGLIDADG meganucleases include, without limitation, the I-Crel meganuclease of SEQ ID NO: 1, the I-Msol meganuclease of SEQ ID NO: 2, and the I-Ceul meganuclease of SEQ ID NO: 3. Wild-type di-LAGLIDADG meganucleases include, without limitation, the I-Dmol meganuclease of SEQ ID NO: 4, the I-Scel meganuclease of SEQ ID NO: 5, and the I-Anil meganuclease of SEQ ID NO: 6.

[0019] Wild-type LAGLIDADG domains include, without limitation, residues 9-151 of the wild-type I-Crel meganuclease of SEQ ID NO: 1; residues 11-162 of the wild-type I-Msol meganuclease of SEQ ID NO: 2; and residues 55-210 of the wild-type I-Ceul meganuclease of SEQ ID NO: 3, residues 9-96 of the wild-type I-Dmol of SEQ ID NO: 4; residues 105-178 of the wild-type I-Dmol of SEQ ID NO: 4; residues 32-123 of the wild-type I-Scel of SEQ ID NO: 5; residues 134-225 of the wild-type I-Scel of SEQ ID NO: 5; residues 4-121 of the wild-type I-Anil of SEQ ID NO: 6; and residues 136-254 of the wild-type I-Anil of SEQ ID NO: 6.

[0020] LAGLIDADG subunits derived from a wild-type LAGLIDADG meganuclease include, without limitation, subunits including a LAGLIDADG domain that has at least 85% sequence identity, or 85%-100% sequence identity, to any one of residues 9-151 of the wild-type I-Crel meganuclease of SEQ ID NO: 1; residues 11-162 of the wild-type I-Msol meganuclease of SEQ ID NO: 2; and residues 55-210 of the wild-type I-Ceul meganuclease of SEQ ID NO: 3, residues 9-96 of the wild-type I-Dmol of SEQ ID NO: 4; residues 105-178 of the wild-type I-Dmol of SEQ ID NO: 4; residues 32-123 of the wild-type I-Scel of SEQ ID NO: 5; residues 134-225 of the wild-type I-Scel of SEQ ID NO: 5; residues 4-121 of the wild-type I-Anil of SEQ ID NO: 6; and residues 136-254 of the wild-type I-Anil of SEQ ID NO: 6.

[0021] LAGLIDADG subunits derived from a wild-type LAGLIDADG meganuclease also include, without limitation, subunits comprising any of the foregoing polypeptide sequences in which one or more amino acid modifications have been included according to the methods of rationally-designing LAGLIDADG meganucleases disclosed in WO 2007/047859 , as well as other non-naturally-occurring meganuclease variants known in the art.

[0022] Disclosed are also recombinant single-chain meganucleases comprising a pair of LAGLIDADG subunits derived from naturally-occurring LAGLIDADG subunits each of which recognizes a wild-type DNA half-site selected from SEQ ID NOs: 7-30.

[0023] Disclosed are also recombinant single-chain meganucleases comprising a pair of LAGLIDADG subunits genetically engineered with respect to DNA-binding specificity, each of which recognizes a DNA half-site that differs by at least one base from a wild-type DNA half-site selected from SEQ ID NOs: 7-30.

[0024] Disclosed are also recombinant single-chain meganucleases comprising a pair of LAGLIDADG subunits in which one subunit is natural and recognizes a wild-type DNA half-site selected SEQ ID NOs: 7-30 and the other is genetically engineered with respect to DNA-binding specificity and recognizes a DNA site that differs by at least one base from a wild-type DNA half-site selected from SEQ ID NOs: 7-30.

[0025] The polypeptide linker joining the LAGLIDADG subunits can be a flexible linker. The linker can include 15-40 residues, 25-31 residues, or any number within those ranges. At least 50%, or 50%-100%, of the residues forming the linker can be polar uncharged residues.

[0026] The polypeptide linker joining the LAGLIDADG subunits can have a stable secondary structure. The stable secondary structure can comprise at least two α-helix structures. The stable secondary structure can comprise from N-terminus to C-terminus a first loop, a first α-helix, a first turn, a second α-helix, and a second loop. The linker can include 23-56 residues, or any number within that range.

[0027] In another aspect are disclosed various methods of use for the single-chain meganucleases described and enabled herein. These methods include producing genetically-modified cells and organisms, treating diseases by gene therapy, treating pathogen infections, and using the recombinant single-chain meganucleases for in vitro applications for diagnostics and research.

[0028] Thus, in one aspect, the invention provides methods for producing a genetically-modified eukaryotic cell including an exogenous sequence of interest inserted in a chromosome, by transfecting the cell with (i) a first nucleic acid sequence encoding a meganuclease of the invention, and (ii) a second nucleic acid sequence including said sequence of interest, wherein the meganuclease produces a cleavage site in the chromosome and the sequence of interest is inserted into the chromosome at the cleavage site either by homologous recombination or non-homologous end-joining.

[0029] Alternatively, in another aspect, the invention provides methods for producing a genetically-modified eukaryotic cell including an exogenous sequence of interest inserted in a chromosome, by introducing a meganuclease protein of the invention into the cell, and transfecting the cell with a nucleic acid including the sequence of interest, wherein the meganuclease produces a cleavage site in the chromosome and the sequence of interest is inserted into the chromosome at the cleavage site either by homologous recombination or non-

homologous end-joining.

[0030] In another aspect, the invention provides methods for producing a genetically-modified eukaryotic cell by disrupting a target sequence in a chromosome, by transfecting the cell with a nucleic acid encoding a meganuclease of the invention, wherein the meganuclease produces a cleavage site in the chromosome and the target sequence is disrupted by non-homologous end-joining at the cleavage site.

[0031] In another aspect, the invention provides methods of producing a non-human genetically-modified organism by producing a genetically-modified eukaryotic cell according to the methods described above, and growing the genetically-modified eukaryotic cell to produce the genetically-modified organism. In these embodiments, the eukaryotic cell is selected from a gamete, a zygote, a blastocyst cell, an embryonic stem cell, and a protoplast cell.

[0032] In another aspect are disclosed methods for treating a disease by gene therapy in a eukaryote, by transfecting at least one cell of the eukaryote with one or more nucleic acids including (i) a first nucleic acid sequence encoding a meganuclease of the invention, and (ii) a second nucleic acid sequence including a sequence of interest, wherein the meganuclease produces a cleavage site in the chromosome and the sequence of interest is inserted into the chromosome by homologous recombination or non-homologous end-joining, and insertion of the sequence of interest provides gene therapy for the disease.

[0033] Alternatively, in another aspect are disclosed methods for treating a disease by gene therapy in a eukaryote, by introducing a meganuclease protein of the invention into at least one cell of the eukaryote, and transfecting the cell with a nucleic acid including a sequence of interest, wherein the meganuclease produces a cleavage site in the chromosome and the sequence of interest is inserted into the chromosome at the cleavage site by homologous recombination or non-homologous end-joining, and insertion of the sequence of interest provides gene therapy for the disease.

[0034] In another aspect are disclosed methods for treating a disease by gene therapy in a eukaryote by disrupting a target sequence in a chromosome of the eukaryotic, by transfecting at least one cell of the eukaryote with a nucleic acid encoding a meganuclease of the invention, wherein the meganuclease produces a cleavage site in the chromosome and the target sequence is disrupted by non-homologous end-joining at the cleavage site, wherein disruption of the target sequence provides the gene therapy for the disease.

[0035] In another aspect are disclosed methods for treating a viral or prokaryotic pathogen infection in a eukaryotic host by disrupting a target sequence in a genome of the pathogen, by transfecting at least one infected cell of the host with a nucleic acid encoding a meganuclease of the invention, wherein the meganuclease produces a cleavage site in the genome and the target sequence is disrupted by either (1) non-homologous end-joining at the cleavage site or (2) by homologous recombination with a second nucleic acid, and wherein disruption of the target sequence provides treatment for the infection.

[0036] Aspects and embodiments of the invention will be apparent to one of ordinary skill in the art based upon the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWING

[0037] Fig. 1 is a diagram of the structural components of one embodiment of a linker of the invention (Linker 9) and N-terminal and C-terminal residues of the endonuclease subunits joined by the liner.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The invention is defined by the appended claims.

1.1 Introduction

[0039] The present invention is based, in part, upon the development of fusion proteins in which a peptide linker covalently joins two heterologous LAGLIDADG meganuclease subunits to form a "single-chain heterodimer meganuclease" in which the subunits function together to preferentially bind to and cleave a non-palindromic DNA recognition site which is a hybrid of the recognition half-sites of the two subunits. In particular, the invention can be used to genetically engineer single-chain meganucleases which recognize non-palindromic DNA sequences that naturally-occurring meganucleases do not recognize.

[0040] This discovery has been used, as is described in detail below, to join mono-LAGLIDADG meganucleases, which naturally function as homodimers, into single-chain meganucleases. Further, the discovery has been used to join mono-LAGLIDADG meganucleases which have been re-engineered with respect to DNA-recognition specificity into single-chain heterodimers which recognize and cleave DNA sequences that are a hybrid of the palindromic sites recognized by the two meganuclease homodimer. The invention provides exemplary peptide linker

sequences for joining LAGLIDADG subunits into single polypeptides. Importantly, the invention provides a general method for the production of linker sequences and the selection of fusion points for linking different LAGLIDADG subunits to produce functional rationally-designed single-chain meganucleases.

[0041] Disclosed are methods that use such meganucleases to produce recombinant nucleic acids, cells and organisms by utilizing the meganucleases to cause recombination of a desired genetic sequence at a limited number of loci within the genome of the organism for, *inter alia,* genetic engineering, gene therapy, treatment of pathogenic infections and cancer, and *in vitro* applications in diagnostics and research.

[0042] As a general matter, disclosed are methods for generating single-chain meganucleases comprising two LAGLIDADG subunits in which the N-terminal subunit is derived from a natural mono-LAGLIDADG meganuclease such as 1-Crel, I-Msol, or I-Ceul or a variant thereof and the C-terminal subunit is derived from either a mono-LAGLIDADG meganuclease or one of the two domains of a di-LAGLIDADG meganuclease such as I-Scel, I-Dmol, or I-Anil. The method is distinct from those described previously (Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62 ; Chevalier et al. (2002), Mol. Cell 10:895-905 ; WO 2003/078619) in that it requires the use of specific and novel linker sequences and fusion points to produce recombinant single-chain meganucleases in which the N-terminal subunit is derived from a mono-LAGLIDADG meganuclease.

[0043] As described in detail below, the method of producing a recombinant single-chain meganuclease includes the use of defined fusion points in the two LAGLIDADG subunits to be joined as well as the use of defined linker sequences to join them into a single polypeptide. In addition, a set of rules is provided for identifying fusion points not explicitly described herein as well as for producing functional linker sequences that are not explicitly described herein.

[0044] Thus, in one aspect, disclosed are methods for producing recombinant single-chain LAGLIDADG meganucleases. In another aspect, disclosed are the recombinant single-chain meganucleases resulting from these methods. In another aspect, disclosed are methods that use such single-chain meganucleases to produce recombinant nucleic acids, cells and organisms in which a desired DNA sequence or genetic locus within the genome of cell or organism is modified by the insertion, deletion, substitution or other manipulation of DNA sequences. In another aspect, disclosed are methods for reducing the survival of pathogens or cancer cells using single-chain meganucleases which have pathogen-specific or cancer-specific recognition sequences.

1.2 References and Definitions

[0045] The patent and scientific literature referred to herein establishes knowledge that is available to those of skill in the art.

[0046] As used herein, the term "meganuclease" refers to an endonuclease that binds double-stranded DNA at a recognition sequence that is greater than 12 base pairs in length. Naturally-occurring meganucleases can be monomeric (e.g., I-Scel) or dimeric *(e.g.,* I-Crel). The term meganuclease, as used herein, can be used to refer to monomeric meganucleases, dimeric meganucleases, to the monomers which associate to form a dimeric meganuclease, or to a recombinant single-chain meganuclease of the invention. The term "homing endonuclease" is synonymous with the term "meganuclease."

[0047] As used herein, the term "LAGLIDADG meganuclease" refers either to meganucleases including a single LAGLIDADG motif, which are naturally dimeric, or to meganucleases including two LAGLIDADG motifs, which are naturally monomeric. The term "mono-LAGLIDADG meganuclease" is used herein to refer to meganucleases including a single LAGLIDADG motif, and the term "di-LAGLIDADG meganuclease" is used herein to refer to meganucleases including two LAGLIDADG motifs, when it is necessary to distinguish between the two. Each of the two structural domains of a di-LAGLIDADG meganuclease which includes a LAGLIDADG motif and has enzymatic activity, and each of the individual monomers of a mono-LAGLIDADG meganuclease, can be referred to as a LAGLIDADG subunit, or simply "subunit".

[0048] As used herein, and in reference to a peptide sequence, "end" refers to the C-terminus and "beginning" refers to the N-terminus. Thus, for example, "the beginning of the LAGLIDADG motif" refers to the N-terminal-most amino acid in the peptide sequence comprising the LAGLIDADG motif whereas "the end of the LAGLIDADG motif" refers to the C-terminal-most amino acid in the peptide sequence comprising the LAGLIDADG motif.

[0049] As used herein, the term "rationally-designed" means non-naturally-occurring and/or genetically engineered. The rationally-designed meganucleases of the invention differ from wild-type or naturally-occurring meganucleases in their amino acid sequence or primary structure, and may also differ in their secondary, tertiary or quaternary structure. In addition, the rationally-designed meganucleases of the invention also differ from wild-type or naturally-occurring meganucleases in recognition sequence-specificity and/or activity.

[0050] As used herein, with respect to a protein, the term "recombinant" means having an altered amino acid sequence as a result of the application of genetic engineering techniques to nucleic acids which encode the protein, and cells or organisms which express the protein. With respect to a nucleic acid, the term "recombinant" means having an altered nucleic acid sequence as a result of the application of genetic engineering techniques. Genetic engineering techniques include, but are not limited to, PCR and DNA cloning technologies; transfection, transformation and other gene transfer technologies; homologous recombination; site-directed mutagenesis; and gene fusion. In accordance with this definition, a protein having an amino acid sequence identical to a naturally-occurring protein, but produced by cloning and

expression in a heterologous host, is not considered recombinant.

[0051] As used herein with respect to recombinant proteins, the term "modification" means any insertion, deletion or substitution of an amino acid residue in the recombinant sequence relative to a reference sequence (e.g., a wild-type).

[0052] As used herein, the term "genetically-modified" refers to a cell or organism in which, or in an ancestor of which, a genomic DNA sequence has been deliberately modified by recombinant technology. As used herein, the term "genetically-modified" encompasses the term "transgenic."

[0053] As used herein, the term "wild-type" refers to any naturally-occurring form of a meganuclease. The term "wild-type" is not intended to mean the most common allelic variant of the enzyme in nature but, rather, any allelic variant found in nature. Wild-type meganucleases are distinguished from recombinant or non-naturally-occurring meganucleases.

[0054] As used herein, the term "recognition sequence half-site" or simply "half site" means a nucleic acid sequence in a double-stranded DNA molecule which is recognized by a monomer of a mono-LAGLIDADG meganuclease or by one LAGLIDADG subunit of a di-LAGLIDADG meganuclease.

[0055] As used herein, the term "recognition sequence" refers to a pair of half-sites which is bound and cleaved by either a mono-LAGLIDADG meganuclease dimer or a di-LAGLIDADG meganuclease monomer. The two half-sites may or may not be separated by base pairs that are not specifically recognized by the enzyme. In the cases of I-CreI, I-MsoI and I-CeuI, the recognition sequence half-site of each monomer spans 9 base pairs, and the two half-sites are separated by four base pairs which are not contacted directly by binding of the enzyme but which constitute the actual cleavage site (which has a 4 base pair overhang). Thus, the combined recognition sequences of the I-CreI, I-MsoI and I-CeuI meganuclease dimers normally span 22 base pairs, including two 9 base pair half-sites flanking a 4 base pair cleavage site. In the case of the I-SceI meganuclease, which is a di-LAGLIDADG meganuclease monomer, the recognition sequence is an approximately 18 bp non-palindromic sequence, and there are no central base pairs which are not specifically recognized. By convention, one of the two strands is referred to as the "sense" strand and the other the "antisense" strand, although neither strand may encode protein.

[0056] As used herein, the term "specificity" means the ability of a meganuclease to recognize and cleave double-stranded DNA molecules only at a particular sequence of base pairs referred to as the recognition sequence, or only at a particular set of recognition sequences. The set of recognition sequences will share certain conserved positions or sequence motifs, but may be degenerate at one or more positions. A highly-specific meganuclease is capable of cleaving only one or a very few recognition sequences. Specificity can be determined in a cleavage assay as described in Example 1. As used herein, a meganuclease has "altered" specificity if it binds to and cleaves a recognition sequence which is not bound to and cleaved by a reference meganuclease (e.g., a wild-type) under physiological conditions, or if the rate of cleavage of a recognition sequence is increased or decreased by a biologically significant amount (e.g., at least 2x, or 2x-10x) relative to a reference meganuclease.

[0057] As used herein, the term "palindromic" refers to a recognition sequence consisting of inverted repeats of identical half-sites. However, the palindromic sequence need not be palindromic with respect to the central base pairs which are not directly contacted by binding of the enzyme (e.g., the four central base pairs of an I-CreI recognition site). In the case of naturally-occurring dimeric meganucleases, palindromic DNA sequences are recognized by homodimers in which the two monomers make contacts with identical half-sites.

[0058] As used herein, the term "pseudo-palindromic" refers to a recognition sequence consisting of inverted repeats of non-identical or imperfectly palindromic half-sites. In addition to central base pairs that are not directly contacted by binding of the enzyme, the pseudo-palindromic sequence can deviate from a palindromic sequence between the two recognition half-sites at 1-3 base pairs at each of the two half-sites. Pseudo-palindromic DNA sequences are typical of the natural DNA sites recognized by wild-type homodimeric meganucleases in which two identical enzyme monomers make contacts with slightly different half-sites.

[0059] As used herein, the term "non-palindromic" refers to a recognition sequence composed of two unrelated half-sites of a meganuclease. In this case, the non-palindromic sequence need not be palindromic with respect to either the central base pairs or 4 or more base pairs at each of the two half-sites. Non-palindromic DNA sequences are recognized by either di-LAGLIDADG meganucleases, highly degenerate mono-LAGLIDADG meganucleases (e.g., I-CeuI) or by heterodimers of mono-LAGLIDADG meganuclease monomers that recognize non-identical half-sites. In the latter case, a non-palindromic recognition sequence may be referred to as a "hybrid sequence" because the heterodimer of two different mono-LAGLIDADG monomers, whether or not they are fused into a single polypeptide, will cleave a recognition sequence comprising one half-site recognized by each monomer. Thus, the heterodimer recognition sequence is a hybrid of the two homodimer recognition sequences.

[0060] As used herein, the term "linker" refers to an exogenous peptide sequence used to join two LAGLIDADG subunits into a single polypeptide. A linker may have a sequence that is found in natural proteins, or may be an artificial sequence that is not found in any natural protein. A linker may be flexible and lacking in secondary structure or may have a propensity to form a specific three-dimensional structure under physiological conditions.

[0061] As used herein, the term "fusion point" refers to the junction between a LAGLIDADG subunit and a linker. Specifically, the "N-terminal fusion point" is the last (C-terminal-most) amino acid of the N-terminal LAGLIDADG subunit prior to the linker sequence and the "C-terminal

fusion point" is the first (N-terminal-most) amino acid of the C-terminal LAGLIDADG subunit following the linker sequence.

[0062] As used herein, the term "single-chain meganuclease" refers to a polypeptide comprising a pair of LAGLIDADG subunits joined by a linker. A single-chain meganuclease has the organization: N-terminal subunit - Linker - C-terminal subunit. A single-chain meganuclease is distinguished from a natural di-LAGLIDADG meganuclease in that the N-terminal subunit must be derived from a mono-LAGLIDADG meganuclease and, therefore, the linker must be exogenous to the N-terminal subunit.

[0063] As used herein, the term "homologous recombination" refers to the natural, cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template (see, *e.g.,* Cahill et al. (2006), Front. Biosci. 11:1958-1976). The homologous DNA sequence may be an endogenous chromosomal sequence or an exogenous nucleic acid that was delivered to the cell. Thus, in some embodiments, a rationally-designed meganuclease is used to cleave a recognition sequence within a target sequence and an exogenous nucleic acid with homology to or substantial sequence similarity with the target sequence is delivered into the cell and used as a template for repair by homologous recombination. The DNA sequence of the exogenous nucleic acid, which may differ significantly from the target sequence, is thereby incorporated into the chromosomal sequence. The process of homologous recombination occurs primarily in eukaryotic organisms. The term "homology" is used herein as equivalent to "sequence similarity" and is not intended to require identity by descent or phylogenetic relatedness.

[0064] As used herein, the term "non-homologous end-joining" refers to the natural, cellular process in which a double-stranded DNA-break is repaired by the direct joining of two non-homologous DNA segments (see, *e.g.* Cahill et al. (2006), Front. Biosci. 11:1958-1976). DNA repair by non-homologous end-joining is error-prone and frequently results in the untemplated addition or deletion of DNA sequences at the site of repair. Thus, in certain embodiments, a rationally-designed meganuclease can be used to produce a double-stranded break at a meganuclease recognition sequence within a target sequence to disrupt a gene (*e.g.,* by introducing base insertions, base deletions, or frameshift mutations) by non-homologous end-joining. In other embodiments, an exogenous nucleic acid lacking homology to or substantial sequence similarity with the target sequence may be captured at the site of a meganuclease-stimulated double-stranded DNA break by non-homologous end-joining (see, *e.g.,* Salomon et al. (1998), EMBD J. 17:6086-6095). The process of non-homologous end-joining occurs in both eukaryotes and prokaryotes such as bacteria.

[0065] As used herein, the term "sequence of interest" means any nucleic acid sequence, whether it codes for a protein, RNA, or regulatory element (*e.g.,* an enhancer, silencer, or promoter sequence), that can be inserted into a genome or used to replace a genomic DNA sequence using a meganuclease protein. Sequences of interest can have heterologous DNA sequences that allow for tagging a protein or RNA that is expressed from the sequence of interest. For instance, a protein can be tagged with tags including, but not limited to, an epitope (*e.g.,* c-myc, FLAG) or other ligand (*e.g.,* poly-His). Furthermore, a sequence of interest can encode a fusion protein, according to techniques known in the art (see, *e.g.,* Ausubel et al., Current Protocols in Molecular Biology, Wiley 1999). In some embodiments, the sequence of interest is flanked by a DNA sequence that is recognized by the recombinant meganuclease for cleavage. Thus, the flanking sequences are cleaved allowing for proper insertion of the sequence of interest into genomic recognition sequences cleaved by the recombinant meganuclease. In some embodiments, the entire sequence of interest is homologous to or has substantial sequence similarity with a target sequence in the genome such that homologous recombination effectively replaces the target sequence with the sequence of interest. In other embodiments, the sequence of interest is flanked by DNA sequences with homology to or substantial sequence similarity with the target sequence such that homologous recombination inserts the sequence of interest within the genome at the locus of the target sequence. In some embodiments, the sequence of interest is substantially identical to the target sequence except for mutations or other modifications in the meganuclease recognition sequence such that the meganuclease can not cleave the target sequence after it has been modified by the sequence of interest.

[0066] As used herein with respect to both amino acid sequences and nucleic acid sequences, the terms "percentage similarity" and "sequence similarity" refer to a measure of the degree of similarity of two sequences based upon an alignment of the sequences which maximizes similarity between aligned amino acid residues or nucleotides, and which is a function of the number of identical or similar residues or nucleotides, the number of total residues or nucleotides, and the presence and length of gaps in the sequence alignment. A variety of algorithms and computer programs are available for determining sequence similarity using standard parameters. As used herein, sequence similarity is measured using the BLASTp program for amino acid sequences and the BLASTn program for nucleic acid sequences, both of which are available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/), and are described in, for example, Altschul et al. (1990), J. Mol. Biol. 215:403 -410; Gish and States (1993), Nature Genet. 3:266-272; Madden et al. (1996), Meth. Enzymol. 266:131-141; Altschul et al. (1997), Nucleic Acids Res. 25:33 89-3402); Zhang et al. (2000), J. Comput. Biol. 7(1-2):203-14. As used herein, percent similarity of two amino acid sequences is the score based upon the following parameters for the BLASTp algorithm: word size = 3; gap opening penalty = -11; gap extension penalty = -1; and scoring matrix = BLOSUM62. As used herein, percent similarity of two nucleic acid sequences is the score based upon the following parameters for the BLASTn algorithm: word size = 11; gap opening penalty = -5; gap extension penalty = -2; match reward = 1; and mismatch penalty = -3.

[0067] As used herein with respect to modifications of two proteins or amino acid sequences, the term "corresponding to" is used to indicate that a specified modification in the first protein is a substitution of the same amino acid residue as in the modification in the second protein, and that the amino acid position of the modification in the first proteins corresponds to or aligns with the amino acid position of the modification in the second protein when the two proteins are subjected to standard sequence alignments (*e.g.,* using the BLASTp program). Thus, the modification of residue "X" to amino acid "A" in the first protein will correspond to the modification of residue "Y" to amino acid "A" in the second protein if residues X and Y correspond to each other in a sequence alignment, and despite the fact that X and Y may be different

numbers.

[0068] As used herein, the recitation of a numerical range for a variable is intended to convey that the invention may be practiced with the variable equal to any of the values within that range. Thus, for a variable which is inherently discrete, the variable can be equal to any integer value within the numerical range, including the end-points of the range. Similarly, for a variable which is inherently continuous, the variable can be equal to any real value within the numerical range, including the end-points of the range. As an example, and without limitation, a variable which is described as having values between 0 and 2 can take the values 0, 1 or 2 if the variable is inherently discrete, and can take the values 0.0, 0.1, 0.01, 0.001, or any other real values ≥ 0 and ≤ 2 if the variable is inherently continuous.

[0069] As used herein, unless specifically indicated otherwise, the word "or" is used in the inclusive sense of "and/or" and not the exclusive sense of "either/or."

2. Single-chain meganucleases derived from LAGLIDADG subunits

[0070] Structural comparisons of natural mono- and di-LAGLIDADG meganucleases reveal that the N-terminal subunits of di-LAGLIDADG meganucleases tend to be smaller than mono-LAGLIDADG monomers. The consequence of this is that, in the case of di-LAGLIDADG meganucleases, the end (C-terminus) of the N-terminal subunit is much closer to the start (N-terminus) of the C-terminal subunit. This means that a relatively short *(e.g., 5-20 amino acid)* linker is sufficient to join the two subunits. In the case of mono-LAGLIDADG meganucleases, the C-terminus of one monomer is generally very far (approximately 48 Å in the case of I-CreI) from the N-terminus of the second monomer. Therefore, fusing a pair of mono-LAGLIDADG meganucleases into a single polypeptide requires a longer *(e.g., >20 amino acid)* peptide linker which can span this distance. An alternative method, in which the N-terminal subunit is truncated at a point spatially closer to the start of the C-terminal subunit has been reported previously (Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62; WO 2003/078619), but produces little if any functional heterodimer, as shown in Example 1 below. An extensive discussion regarding the difficulty associated with producing a functional single-chain meganuclease derived from I-CreI can be found in Fajardo-Sanchez et al. (2008), Nucleic Acids Res. 36:2163-2173.

2.1 Fusion points for I-CreI

[0071] A series of truncation mutants were made in which either wild-type I-CreI or an engineered variant of I-CreI which had been altered with respect to its DNA cleavage site preference (designated "CCR2", SEQ ID NO: 31; see WO 2007/047859) were terminated prior to the natural C-terminal amino acid, Pro 163 (Table 1). The mutant homodimers were expressed in *E. coli*, purified, and incubated with either the wild-type recognition sequence (SEQ ID NOs: 34-35) or the CCR2 recognition sequence (SEQ ID NOs: 32-33) to test for cleavage activity.

TABLE 1

I-CreI Truncation Mutants		
C-terminal amino acid	Wild-type activity	CCR2 activity
Asp-153	+	+
Val-151	+	+
Val-148	+	-
Arg-141	-	-
Asn-136	-	-
Val-129	-	-
Ile-109	-	-
Leu-95	-	-

[0072] Wild-type I-CreI was found to be active when truncated at residue 148 or further C-terminal residues, but inactive when truncated at residue 141 or further N-terminal residues. Therefore, at least some of residues 141 through 147, or conservative substitutions of those residues, are required for wild-type activity. Similarly, CCR2 was found to be active when truncated at residue 151 or further C-terminal residues, but inactive when terminated at residue 148 or further N-terminal residues. Therefore, at least some of residues 148 through 150, or conservative substitutions of those residues, are required for CCR2 activity. The difference between the wild-type I-CreI and the rationally-designed CCR2 meganuclease is presumably due to a reduction in the structural stability of the CCR2 meganuclease such that it is more sensitive to further destabilization by a premature C-terminal truncation. These truncation results are consistent with a publication from Prieto *et al.* in which it was found that the C-terminal loop of I-CreI (amino acids 138-142) is essential for cleavage activity (Prieto et al. (2007), Nucl. Acids Res. 35:3262-3271). Taken together, these results indicate that some residues near the C-terminus of I-CreI are essential for DNA-binding and/or catalytic activity and methods for single-chain meganuclease production that truncate an I-CreI subunit prior to approximately residue 142 *(e.g.,* Epinat et al. (2003), Nucl. Acids Res. 31: 2952-62; WO 2003/078619) are unlikely to yield a single-chain meganuclease in which both LAGLIDADG subunits are catalytically active.

8

[0073] Therefore, in accordance with the present invention, the N-terminal fusion point (*i.e.*, between the N-terminal I-CreI subunit and the linker) should lie at or C-terminal to residue 142 of the N-terminal subunit, including any of positions 142-151, or any position C-terminal to residue 151. Residues 154-163 of I-CreI are unstructured (Jurica et al. (1998), Mol. Cell 2:469-476) and, therefore, inclusion of these residues will increase the flexibility and, possibly, structural instability of the resultant single-chain meganuclease. Conversely, if it is determined that less flexibility and greater structural stability are desired or required, fusion points at residues 142-153 can be chosen.

[0074] When the C-terminal LAGLIDADG subunit of a single-chain meganuclease is derived from I-CreI, the C-terminal fusion point of the linker will be toward the N-terminus of the I-CreI sequence. Residues 7,8 and 9 are of particular interest as C-terminal fusion points in I-CreI because these residues (1) are structurally conserved among LAGLIDADG meganuclease family members and, therefore, may provide greater compatibility in forming heterodimers with other LAGLIDADG family members, and (2) initiate an alpha-helix containing the conserved LAGLIDADG motif that is involved in catalytic function. However, fusion points N-terminal to residue 7, including any of residues 1-6, can also be employed in accordance with the invention.

[0075] The following I-CreI N-terminal and C-terminal fusion points were chosen for further experimentation, but should not be regarded as limiting the scope of the invention:

TABLE 2

I-CreI Fusion Points	
N-terminal fusion point	C-terminal fusion point
Val-151	Lys-7
Leu-152	Asp-8
Asp-153	Phe-9

2.2 Linkers for single-chain meganucleases derived from I-CreI

[0076] For the purpose of linking a pair of I-CreI monomers into a single polypeptide, two general classes of linker were considered: (1) an unstructured linker lacking secondary structure; and (2) a structured linker having secondary structure. Examples of unstructured linkers are well known in the art, and include artificial sequences with high Gly and Ser content, or repeats. Structured linkers are also well known in the art, and include those designed using basic principles of protein folding (*e.g.*, Aurora and Rose (1998), Protein Sci. 7:21-38; Fersht, Structure and Mechanism in Protein Science, W.H. Freeman 1998).

[0077] The invention was validated using a pair of rationally-designed I-CreI monomers called "LAM1" (SEQ ID NO: 36) and "LAM2" (SEQ ID NO: 37). These rationally-designed endonucleases were produced using the methods described in WO 2007/047859 and they are characterized therein. As will be apparent to those of skill in the art, however, the LAM1 and LAM2 monomers are merely exemplary of the many monomers which can be employed, including wild-type mono-LAGLIDADG subunits, N-terminally and/or C-terminally truncated wild-type mono-LAGLIDADG subunits, N-terminally and/or C-terminally truncated wild-type di-LAGLIDADG subunits, and rationally designed modifications of any of the foregoing.

[0078] One exemplary monomer, LAM1, differs by 7 amino acids from wild-type I-CreI and recognizes the half site:

5'-TGCGGTGTC-3' (SEQ ID NO: 38)

3'-ACGCCACAG-5' (SEQ ID NO: 39)

[0079] Thus, the LAM 1 homodimer recognizes the palindromic recognition sequence (where each N is unconstrained):

5'-TGCGGTGTCNNNNGACACCGCA-3' (SEQ ID NO: 40)

3'-ACGCCACAGNNNNCTGTGGCGT-5' (SEQ ID NO: 41)

[0080] The other exemplary monomer, LAM2, differs by 5 amino acids from wild-type I-CreI and recognizes the half-site:

5'-CAGGCTGTC-3' (SEQ ID NO: 42)

3'-GTCCGACAG-5' (SEQ ID NO: 43)

[0081] Thus, the LAM2 homodimer recognizes the palindromic recognition sequence (where each N is unconstrained):

5'-CAGGCTGTCNNNNGACAGCCTG-3' (SEQ ID NO: 44)

3'-GTCCGACAGNNNNCTGTCGGAC-5' (SEQ ID NO: 45)

[0082] A heterodimer comprising one LAM1 monomer and one LAM2 monomer ("LAM1/LAM2 heterodimer") thus recognizes the hybrid recognition sequence:

5'-TGCGGTGTCNNNNGACAGCCTG-3' (SEQ ID NO: 40)

3'-ACGCCACAGNNNNCTGTCGGAC-5' (SEQ ID NO: 41)

2.2.1 Flexible linkers for single-chain meganucleases

[0083] A variety of highly-flexible peptide linkers are known in the art and can be used in accordance with the invention. For example, and without limitation, peptide linkers comprising repeating Gly-Ser-Ser units are known to be unstructured and flexible (Fersht, Structure and Mechanism in Protein Science, W.H. Freeman 1998). Linkers with this and similar compositions are frequently used to fuse protein domains together (*e.g.*, single-chain antibodies (Mack et al. (1995), Proc. Nat. Acad. Sci. 92:7021-7025); growth factor receptors (Ueda et al. (2000), J. Immunol. Methods 241:159-170); enzymes (Brodelius *et al.* (2002), 269:3570-3577); and DNA-binding and nuclease domains (Kim et al. (1996), Proc. Nat. Acad. Sci. 93:1156-1160).

[0084] As a general matter, the flexible linker can include any polypeptide sequence which does not form stable secondary structures under physiological conditions. In some embodiments, the linkers include a high percentage (*e.g.*, > 50%, 60%, 70%, 80% or 90%, or generally, 50%-100%) of polar uncharged residues (*i.e.*, Gly, Ser, Cys, Asn, Gln, Tyr, Thr). In addition, in some embodiments, the linkers include a low percentage of large hydrophobic residues (*i.e.*, Phe, Trp, Met). The linkers may include repeats of varying lengths (*e.g.*, $(SG)_n$, $(GSS)_n$, $(SGGS)_n$), may include random sequences, or may include combinations of the two.

[0085] Thus, in accordance with the invention, a set of single-chain fusions between LAM1 and LAM2 were produced in which a highly-flexible peptide linker covalently joined the N-terminal (LAM1) subunit to the C-terminal (LAM2) subunit using Val-151 or Asp-153 as the N-terminal fusion point and Phe-9 as the C-terminal fusion point. The single-chain proteins were expressed in *E. coli*, purified, and tested for the ability to cleave a hybrid DNA site comprising one LAM1 half-site and one LAM2 half-site (SEQ ID NOs: 46 and 47). Cleavage activity was rated on a four point scale: - no detectable activity; + minimal activity; ++ medium activity; +++ activity comparable to the LAM1/LAM2 heterodimer produced by co-expression of the two monomers in *E. coli* prior to endonuclease purification. The proteins were also evaluated by SDS-PAGE to determine the extent to which the linker region was proteolyzed during expression or purification to release the two subunits.

TABLE 3

Single-Chain I-Crel Meganucleases with Gly-Ser Linkers					
Linker number	N-term. fusion pt.	C-term. fusion pt.	Linker sequence	Activity	Linker proteolysis
1	Val-151	Phe-9	$(GSS)_7G$	-	-
2	Val-151	Phe-9	$(GSS)_8G$	-	-
3	Val-151	Phe-9	$(GSS)_9G$	+	+
4	Val-151	Phe-9	$(GSS)_{10}G$	ND	+++
5	Val-151	Phe-9	$(GSS)_{11}G$	ND	+++
6	Val-151	Phe-9	$(GSS)_9GG$	+	+
7	Val-151	Phe-9	$(GSS)_9GSG$	+	+
8	Asp-153	Phe-9	$(GSS)_9G$	+	+

[0086] The results indicated that flexible linkers, such as the Gly-Ser linkers in Table 3, are suitable for single-chain meganuclease production provided that the length is appropriate (see also Example 2). For example, referring to Table 3, single-chain meganucleases including linkers 1 and 2, comprising 22 and 25 total amino acids, respectively, did not exhibit any detectable cleavage activity with the fusion points tested. SDS-PAGE indicated that these meganucleases were intact and were not degraded by proteases, leading to the conclusion that these single-chain meganucleases were structurally stable but functionally constrained by linkers that were too short to allow the individual LAGLIDADG subunits to adopt the necessary conformation for DNA binding and/or catalysis. Linkers 3, 6, 7, and 8, comprising 28, 29, 30, and 28 amino acids, respectively, all exhibited low levels of cleavage activity. SDS-PAGE indicated that a small amount (5%-10%) of

each was proteolyzed into individual subunits while the majority had a molecular weight corresponding to the full-length single-chain meganuclease (~40 kilodaltons). Numbers 3 and 8 have the same linker sequence but N-terminal fusion points at Val-151 and Asp-153, respectively. Both single-chain meganucleases exhibited comparable levels of activity, indicating that the precise fusion point is not critical in this instance. Finally, linkers 4 and 5, comprising 31 and 34 amino acids, respectively, yielded no detectable single-chain meganuclease when purified from *E. coli*. These linkers were completely proteolyzed to the individual LAM1 and/or LAM2 subunits as detected by SDS-PAGE and, therefore, the cleavage activity of these meganucleases was not investigated further.

[0087] These results led us to conclude that Gly-Ser linkers are acceptable for the production of single-chain meganucleases based upon the LAGLIDADG subunit of the mono-LAGLIDADG meganuclease I-CreI and the particular fusion points employed, provided that the linkers are greater than 25 and less than 31 amino acids in length. For I-CreI-based single-chain meganucleases with these fusion points, shorter linkers prevent catalysis while longer linkers are unstable and prone to clipping by proteases.

[0088] The effects of varying the fusions points on the acceptable linker lengths can be determined empirically by routine experimentation and/or predicted based upon three-dimensional modeling of the protein structures. Significantly, as a fusion point is moved either N-terminally or C-terminally, it may move either closer or farther from the other fusion point depending upon the secondary and tertiary structure of the protein near the fusion point. Thus, for example, moving the N-terminal fusion point in the C-terminal direction (*e.g.*, from residue 150 to residue 155 for an N-terminal subunit) does not necessarily result in the N-terminal fusion point being physically closer to the C-terminal fusion point because, for example, the N-terminal residues in that region may be part of a secondary/tertiary structure that is pointing either towards or away from the C-terminal fusion point. Thus, moving an N-terminal fusion point in either the N-terminal or C-terminal direction, or moving a C-terminal fusion point in either the N-terminal or C-terminal direction, can result in a shift in the range of acceptable linker lengths toward either longer or shorter linkers. That shift, however, is readily determined, as shown by the experiments reported herein, by routine experimentation and/or three-dimensional modeling.

[0089] Thus, in some embodiments, useful flexible linkers have lengths of greater than 25 residues and less than 31 residues (including all values in between), as shown in Table 3 for a single-chain meganuclease based on two I-CreI LAGLIDADG subunits. In other embodiments, however, employing different LAGLIDADG subunits and/or different fusion points, useful flexible linkers can have lengths greater than 15 and less than 40 residues (including all values in between), provided that the linkers are not extensively proteolyzed and that the single-chain meganuclease retains DNA-binding and cleavage activity as determined by the simple assays described herein.

2.2.2 Designed, structured linkers for single-chain meganucleases

[0090] In an effort to produce single-chain I-CreI-based meganucleases with nuclease activity comparable to the natural dimeric enzyme which are both stable enough for long-term storage and resistant to proteolysis, linkers having stable secondary structures can be used to covalently join subunits. A search of the Protein Databank (www.rcsb.org) did not reveal any structurally-characterized LAGLIDADG proteins with linkers suitable for spanning the great distance (approx. 48 Å) between the identified N-and C-terminal fusion points in I-CreI. Therefore, known first principles governing protein structure *(e.g.*, Aurora and Rose (1998), Protein Sci. 7:21-38; Fersht, Structure and Mechanism in Protein Science, W.H. Freeman 1998) were employed to produce a set of linkers expected to have structural elements suitable for joining the two subunits. Specifically, it was postulated that a suitable linker would comprise (listed from N-terminal fusion point to C-terminal fusion point):

(1) Loop 1. This structural element starts at the N-terminal fusion point and reverses the direction of the peptide chain back on itself (a 180° turn). The sequence can be 3-8 amino acids and can include at least one glycine residue or, in some embodiments, 2-3 glycines. This structural element can be stabilized by introducing a "C-capping" motif to terminate the C-terminal α-helix of I-CreI and initiate the subsequent turn. The helix cap motif is typically described as beginning with a hydrophobic amino acid in the final turn of the helix (Aurora and Rose (1998), Protein Sci. 7:21-38). The C-cap can take any of the forms listed in Table 4:

TABLE 4

C-capping Motifs	
Number	Motif
1	h_1xpx-Gh
2	h_1xpx-nxhx
3	h_1xpx-nxph
4	h_1xxx-Gphx
5	h_1xxx-Gpph
6	h_1xxx-Pppph
7	h_1xxx-Ppph

where h = a hydrophobic amino acid (Ala, Val, Leu, Ile, Met, Phe, Trp, Thr, or Cys); p = a polar amino acid (Gly, Ser, Thr, Asn, Gln, Asp, Glu, Lys, Arg); n = a non-β-branched amino acid (not Val, Ile, Thr, or Pro); x = any amino acid from the h or p group; G = glycine; and P = proline. Note that Thr appears in both groups h and p because its side chain has both hydrophobic (methyl group) and polar (hydroxyl) functional groups. The hyphen designates the end of the α-helix and h_1 is a hydrophobic amino acid in the final turn of the helix (*i.e.*, a hydrophobic

amino acid 0-4 amino acids prior to the N-terminal fusion point). In the case of I-CreI, h_1 is typically Val-151 or Leu-152. Thus, an example of motif 7 is the sequence $V_{151}L_{152}D_{153}$S-PGSV (see, for example, Table 6, Linker 9).

(2) α-helix 1. Following Loop 1, this first α-helix in the linker is designed to run anti-parallel to the C-terminal helix in I-CreI (amino acids 144-153) on the outside face of the protein for a distance of approximately 30 Å. This segment should be 10-20 amino acids in length, should not contain any glycine or proline amino acids outside of the N- and C-capping motifs (below), and alternate hydrophobic and polar amino acids with 3-4 amino acid periodicity so as to bury one face of the helix (the hydrophobic face) against the surface of the N-terminal I-CreI subunit while exposing the other face to solvent. The helix could, for example, take the form pphpphhpphpp where p is any polar amino acid and h is any hydrophobic amino acid but neither is glycine or proline such as the sequence SQASSAASSASS (see, for example, Table 6, Linker 9). Numerous algorithms are available to determine the helical propensity of a peptide sequence (e.g., BMERC-PSA, http://bmerc-www.bu.edu/psa/; NNPREDICT, http://alexander.compbio.ucsf.edu/~nomi/nnpredict.html; PredictProtein, http://www.predictprotein.org) and any of these can be used to produce a sequence of the appropriate length that can be expected to adopt α-helical secondary structure. Alternatively, this helix sequence could be derived from a peptide sequence known to adopt α-helical secondary structure in an existing natural or designed protein. Numerous examples of such peptide sequences are available in the Protein Databank (www.rcsb.org).

In addition, it may be desirable to initiate the α-helix with an N-capping motif to stabilize its structure (Aurora and Rose (1998), Protein Sci. 7:21-38). This motif spans the loop - α-helix junction and typically has one of the forms shown in Table 5:

TABLE 5

N-capping Motifs	
Number	**Motif**
1	h-xpxhx
2	h-xxpph
3	hp-xpxhx
4	hp-xxpph
5	hpp-xpxhx
6	hpp-xxpph

where the designations are as in Table 4 above and the hyphen represents the junction between the loop and the helix. An example of motif number 2 is the sequence L-SPSQA (see, for example, Table 6, Linker 9).

(3) Turn 1. Following α-helix 1, a short, flexible peptide sequence is introduced to turn the overall orientation of the peptide chain by approximately 90° relative to the orientation of α-helix 1. This sequence can be 3-8 amino acids in length and can contain 1 or, in some embodiments, 2-3 glycines. This sequence can also contain a C-cap such as one of the motifs in Table 4 to stabilize α-helix 1 and initiate the turn. An example is the sequence ASSS-PGSGI (see, for example, Table 6, Linker 9) which conforms to C-capping motif number 6. In this case, the sequence ASSS is the final turn of α-helix 1 while the sequence PGSGI is Turn 1.

(4) α-helix 2. This helix follows Turn 1 and is designed to lie at the surface of I-CreI in a groove created at the interface between the LAGLIDADG subunits. The surface of this groove comprises primarily amino acids 94-100 and 134-139 of the N-terminal subunit and amino acids 48-61 of the C-terminal subunit.

α-helix 2 can be designed to be shorter than α-helix 1 and can comprise 1-3 turns of the helix (4-12 amino acids). α-helix 2 can have the same overall amino acid composition as α-helix 1 and can also be stabilized by the addition of an N-capping motif of Table 5. The sequence I-SEALR is an example (see, for example, Table 6, Linker 9) that conforms to N-capping motif number 1. Linker 9 incorporates a relatively short α-helix 2 comprising the sequence SEALRA which is expected to make approximately two turns. Experiments with different linker α-helix 2 sequences have demonstrated the importance of helical register in this region of the linker. The addition of a single amino acid (e.g., A, Linker 11), two amino acids (e.g., AS, Linker 12), or three amino acids (e.g., ASS, Linker 13) prior to the termination of α-helix 2 with a glycine amino acid can result in single-chain I-CreI proteins that are unstable and precipitate within moments of purification from E. coli (Table 6). In contrast, the addition of four amino acids (e.g., ASSA, linker 14), which is expected to make one full additional turn and restore the helical register to that of Linker 1 is stable and active.

(5) Loop 2. This loop terminates α-helix 2 and turns the peptide chain back on itself to join with the C-terminal I-CreI subunit at the C-terminal fusion point. As with Loop 1, this sequence can be 3-8 amino acids in length and can contain one or more glycines. It can also contain a C-capping motif from Table 4 to stabilize α-helix 2. For example, the sequence ALRA-GA from Linker 9 conforms to C-capping motif number 1. In addition, this segment can begin an N-cap on the N-terminal α-helix (amino acids 7-20) of the C-terminal I-CreI subunit. For example the sequence T-KSK$_7$E$_8$F$_9$ from Linker 9 conforms to N-capping motif number 2. In this instance, the C-terminal fusion point is Lys-7. In other cases, the fusion point can be moved further into the second subunit (for example to amino acids 8 or 9), optionally with the addition of 1-2 amino acids to Loop 2 to compensate for the change in helical register as the C-terminal fusion point is moved. For example, linkers 15-23 in Table 6 below have Glu-8 as the C-terminal fusion point and all have an additional amino acid in Loop 2 relative to Linkers 1-6.

[0091] Employing the principles described above, the set of linkers outlined in Table 6 were developed. A set of single-chain I-CreI meganucleases incorporating the linkers between LAM1 and LAM2 subunits was constructed and each was tested for activity against the LAM1/LAM2 hybrid recognition sequence. In all cases, the N-terminal fusion point was Asp-153 of LAM1 and the C-terminal fusion point was either Lys-7 or Glu-8 (denoted in the "CFP" column) of LAM2. Cleavage activity was rated on a four point scale: - no detectable activity; + minimal activity; ++ medium activity; +++ activity comparable to the LAM1/LAM2 heterodimer produced by co-expression of the two monomers in E. coli prior to endonuclease purification. Immediately following purification, the single-chain meganucleases were centrifuged (2100g for 10 minutes) to pellet precipitated protein (indicative of structural instability) and the amount of precipitate (ppt) observed was scored: - no

precipitate; + slight precipitate; ++ significant precipitate. Those protein samples that precipitated to a significant degree could not be assayed for cleavage activity.

TABLE 6

#	CFP	Linker Sequence	Activity	ppt
		Linkers for Single-Chain I-CreI		
9	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKS	+++	-
10	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAGGATKS	+++	-
11	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAAGGATKS	ND	++
12	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAASGGATKS	ND	++
13	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAASSGGATKS	ND	++
14	K7	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAASSAGGATKS	+++	-
15	E8	SLPGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKEF	++	+
16	E8	SLPGSVGGLSPSQASSAASSASSSPGSGTSEAPRAGATKEF	++	-
17	E8	SLPGSVGGLSPSQASSAASSASSSPGSGTSEATRAGATKEF	++	+
18	E8	SLPGSLGGLSPSQASSAASSASSSPGSGPSEALRAGATKEF	++	+
19	E8	SLPGSVGGLSPSQASSAASSASSSPGSGVSEASRAGATKEF	++	+
20	E8	SLPGSVGGLSPSQASSAASSASSSPGSGLSEALRAGATKEF	++	+
21	E8	SLPGSLGGISPSQASSAASSASSSPGSGSSEASRAGATKEF	++	-
22	E8	SPGSVGGVSPSQASSAASSASSSPGSGISEATRAGATKEF	++	-
23	E8	SLPGSLGGVSPSQASSAASSPGSGTSEAPRAGATKEF	ND	++
24	E8	SLPGSVGGLSPSQASSAASSPGSGISEAIRAGATKEF	++	-
25	E8	SLPGSLGGVSPSQASSAASSASSAASSPGSGASEASRAGATKEF	++	-

Single-chain meganucleases each of these linkers except for 11-13 and 23 (which were not investigated) ran as a single band of the desired molecular weight (~40 kilodaltons) on an SDS-PAGE gel, indicative of resistance to proteolytic cleavage of the linker sequence. In at least one case (Linker 9), the single-chain LAM meganuclease could be stored at 4°C in excess of 4 weeks without any evidence of degradation or loss of cleavage activity. Moreover, a number of single-chain LAM endonucleases (9, 10, and 14) cleaved the hybrid LAM1/LAM2 recognition sequence with efficiency comparable to the purified LAM1/LAM2 heterodimer, indicating that fusing I-CreI subunits using these linkers does not significantly impair endonuclease activity (see Example 2).

[0092] In stark contrast to the purified LAM1/LAM2 heterodimer (which is, in fact, a mixture of homo- and heterodimers), the single-chain LAM meganucleases incorporating the linkers in Table 6 cleave the hybrid site much more efficiently than either of the palindromic sequences (see Example 2). The palindromic sequences are typically cut with <5% efficiency relative to the hybrid site. This unintended cleavage of the palindromic DNA sites could be due to (1) homo-dimerization of LAM1 or LAM2 subunits from a pair of different single-chain proteins, (2) sequential nicking of both strands of the palindromic sequence by a single subunit (LAM1 or LAM2) within the single-chain meganuclease, or (3) minute amounts of homodimeric LAM1 or LAM2 that form following proteolytic cleavage of the single-chain meganuclease into its individual subunits (although SDS-PAGE results make this latter explanation unlikely). Although the single-chain I-CreI meganucleases maintain some activity toward palindromic DNA sites, the activity is so disproportionately skewed in favor of the hybrid site that this approach represents a very significant improvement over existing methods.

3. Single-chain meganucleases derived from I-Msol

[0093] I-Msol is a close structural homolog of I-CreI and similar methods have been presented for redesigning the DNA-binding specificity of this meganuclease (WO 2007/047859). The method presented above for the production of a single-chain I-CreI meganuclease can be directly applied to I-Msol. Amino acids Phe-160, Leu-161, and Lys-162 of I-Msol are structurally homologous to, respectively, Val-151, Leu-152, and Asp-153 of I-CreI. These amino acids, therefore, can be selected as the N-terminal fusion points for I-Msol. In addition, The X-ray crystal structure of I-Msol reveals that amino acids 161-166 naturally act as a C-cap and initiate a turn at the C-terminus of the protein which reverses the direction of the peptide chain. Thus, Ile-66 can be selected as the N-terminal fusion point provided that the linker is shortened at its N-terminus to remove the C-cap portion of Loop 1. Pro-9, Thr-10, and Glu-11 of I-Msol are structurally homologous to, respectively, Lys-7, Glu-8, and Phe-9 of I-CreI and can be selected as C-terminal fusion points for I-Msol (Table 7). In addition, because the sequence $L_7Q_8P_9T_{10}E_{11}A_{12}$ of I-Msol forms a natural N-cap (motif 2 from Table 5), Leu-7 can be included as a fusion point.

TABLE 7

I-Msol Fusion Points	
N-terminal fusion points	**C-terminal fusion points**
Phe-160	Leu-7

I-Msol Fusion Points	
N-terminal fusion points	**C-terminal fusion points**
Leu-161	Pro-9
Lys-162	Thr-10
Ile-166	Glu-11

[0094] Any of the linkers in Tables 3 or 6 can be used for the production of single-chain I-Msol endonucleases. For example, Linker 9 from Table 6 may be used to join a pair of I-Msol subunits into a functional single-chain meganuclease using Lys-162 and Pro-9 as fusion points. In one embodiment, Pro-9 is changed to a different amino acid (e.g., alanine or glycine) because proline is structurally constraining. This is analogous to selecting Thr-10 as the C-terminal fusion point and adding an additional amino acid to the C-terminus of the linkers listed in Tables 3 or 6. For example Linkers 26 and 27 from Table 8 are identical to Linker 9 from Table 6 except for the addition of a single amino acid at the C-terminus to account for a change in C-terminal fusion point from Pro-9 (structurally homologous to I-Crel Lys-7) to Thr-10 (structurally homologous to I-Crel Glu-8).

[0095] In another embodiment, as described in Example 4, a single-chain meganuclease derived from I-Mso can also be successfully produced using a linker sequence selected from Linker 28-30 from Table 8 in which I-166 is selected as the N-terminal fusion point and Leu-7 is selected as the C-terminal fusion point. Because I-166 is selected as the N-terminal fusion point, the C-cap portion of Loop 1 (corresponding to the first 6 amino acids of each of the linkers from Table 6) can be removed. In addition, α-helix 1 of Linkers 28-30 are lengthened by 3 amino acids (AAS, underlined in Table 8) relative to the linkers listed in Table 6, corresponding to one additional turn of the helix. Using Linkers 28-30 and the specified fusion points, it is possible to produce protease-resistant, high-activity single-chain meganucleases comprising a pair of I-Mso-derived subunits (see Example 4).

TABLE 8

Linkers for Single-Chain I-Msol					
#	NFP	CFP	Linker Sequence	Activity	ppt
26	K162	T10	PGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKSA	++	-
27	K162	T10	PGSVGGLSPSQASSAASSASSSPGSGISEALRAGATKSG	++	-
28	I166	L7	GGASPSQASSAASSASS<u>AAS</u>SPGSGISEALRAASSLASKPGST	+++	-
29	I166	L7	GGASPSQASSAASSASS<u>AAS</u>SPGSGISEALRAASSPGST	+++	-
30	I166	L7	GGASPSQASSAASSASS<u>AAS</u>SPGSGPSEALRAASSFASKPGST	+++	-

4. Single-chain meganucleases derived from I-Ceul

[0096] I-Ceul is a close structural homolog of I-Crel and similar methods have been presented for redesigning the DNA-binding specificity of this meganuclease (WO 2007/047859). The method presented above for the production of a single-chain I-Crel meganuclease can be directly applied to I-Ceul. Amino acids Ala-210, Arg-211, and Asn-212 ofl-Ceul are structurally homologous to, respectively, Val-151, Leu-152, and Asp-153 of I-Crel. These amino acids, therefore, can be selected as the N-terminal fusion points for I-Ceul. Ser-53, Glu-54, and Ser-55 of I-Ceul are structurally homologous to, respectively, Lys-7, Glu-8, and Phe-9 of I-Crel and can be selected as C-terminal fusion points for I-Ceul (Table 9).

TABLE 9

I-Ceul Fusion Points	
N-terminal fusion points	**C-terminal fusion points**
Ala-210	Ser-53
Arg-211	Glu-54
Asn-212	Ser-55

[0097] Any of the linkers in Tables 3 or 6 can be effective for the production of single-chain I-Ceul endonucleases. For example, I-Ceul subunits can be joined by Linker 9 from Table 6 using Asn-212 as the N-terminal fusion point and Ser-53 as the C-terminal fusion point.

[0098] The C-terminal fusion points selected for I-Ceul result in the removal of amino acids 1-52 from the C-terminal I-Ceul subunit. Structural analyses (Spiegel et al. (2006), Structure 14:869-880) reveal that these amino acids form a structured domain that rests on the surface ofl-Ceul and buries a substantial amount of hydrophobic surface area contributed by amino acids 94-123. It is possible, therefore, that removing this N-terminal domain will destabilize the C-terminal I-Ceul subunit in the single-chain meganuclease. To mitigate this possibility, the hydrophobic amino acids that would be exposed by the removal of this N-terminal domain can be mutated to polar amino acids

(e.g., non-β-branched, hydrophobic amino acids can be mutated to Ser while β-branched, hydrophobic amino acids can be mutated to Thr). For example, Leu-101, Tyr-102, Leu-105, Ala-121, and Leu-123 can be mutated to Ser while Val-95, Val-98, and Ile-113 can be mutated to Thr.

[0099] Alternatively, the N-terminal domain of the C-terminal I-CeuI subunit can be left largely intact and joined to the N-terminal subunit via a truncated linker. This can be accomplished using Lys-7, Pro-8, Gly-9, or Glu-10 as a C-terminal fusion point. The linker can be a flexible Gly-Ser linker (e.g., Linker 3 from Table 3) truncated by approximately 50% of its length (i.e., $(GSS)_4G$ or $(GSS)_5G$). Alternatively, the linker can be any of the linkers from Table 6 truncated within Turn 1. Thus, using Linker 9 from Table 6 as an example, a single-chain I-CeuI meganuclease can be made with the following composition:

N-term. subunit N_{212}-SLPGSVGGLSPSQASSAASSASSSPGS-G_9 C-term. subunit

5. Single-chain meganucleases derived from two different LAGLIDADG family members

[0100] This invention also enables the production of single-chain meganucleases in which each of the subunits is derived from a different natural LAGLIDADG domain. "Different," as used in this description, refers to LAGLIDADG subunits that are not derived from the same natural LAGLIDADG family member. Thus, as used in this description, rationally-designed LAGLIDADG subunits from the same family member (e.g., two I-CreI subunits that have been genetically engineered with respect to DNA cleavage specificity) are not considered "different". Specifically, the invention enables the production of single-chain meganucleases comprising an N-terminal subunit derived from a mono-LAGLIDADG meganuclease (e.g., I-CreI, I-MsoI, or I-CeuI) linked to a C-terminal subunit derived from a different mono-LAGLIDADG meganuclease or either of the two LAGLIDADG domains from a di-LAGLIDADG meganuclease. For example, a single-chain meganuclease can be produced comprising an N-terminal I-CreI subunit, which may or may not have been rationally-designed with regard to DNA recognition site specificity, linked to a C-terminal I-MsoI subunit which also may or may not have been rationally-designed with regard to DNA recognition site specificity.

[0101] In the cases of I-CreI, I-MsoI, and I-CeuI, the desirable fusion points and linkers are as described above. For example, a single-chain I-CreI to I-MsoI fusion can be produced using Linker 9 from Table 6 to join I-CreI Asp-153 to I-MsoI Thr-10. Table 9 lists potential C-terminal fusion points for individual LAGLIDADG domains from the di-LAGLIDADG meganucleases I-SceI, I-DmoI, and I-AniI.

TABLE 10

C-terminal Fusion Points for di-LAGLIDADG Meganuclease Subunits					
I-SceI N-terminal (31-123)	I-SceI C-terminal (132-225)	I-AniI N-terminal (3-125)	I-AniI C-terminal (135-254)	I-DmoI N-terminal (8-98)	I-DmoI C-terminal (104-178)
I-31	Y-132	D3	S-135	S-8	R-104
E-32	L-133	L4	Y-136	G-9	E-105
Q-33	T-134	Y6	F-137	I-10	Q-106

The fusion points listed in Tables 7, 9 and 10 are based on structure comparisons between the meganuclease in question and I-CreI in which amino acid positions which are structurally homologous to the I-CreI fusion points were identified. Fusion points can also be identified in LAGLIDADG subunits which have not been structurally characterized using protein sequence alignments to I-CreI. This is particularly true of C-terminal fusion points which can be readily identified in any LAGLIDADG subunit based upon the location of the conserved LAGLIDADG motif. The amino acids which are 4-6 residues N-terminal of the start of the LAGLIDADG motif are acceptable C-terminal fusion points.

[0102] Because the dimerization interfaces between subunits from different LAGLIDADG endonucleases vary, the subunits may not associate into functional "heterodimers" despite being covalently joined as a single polypeptide. To promote association, the interface between the two subunits can be rationally-designed, as described in WO 2007/047859. At its simplest, this involves substituting interface residues from one subunit onto another. For example, I-CreI and I-MsoI differ in the interface region primarily at I-CreI Glu-8 (which is a Thr in the homologous position of I-MsoI, amino acid 10) and Leu-11 (which is an Ala in the homologous position of I-MsoI, amino acid 13). Thus, I-CreI and I-MsoI subunits can be made to interact effectively by changing Glu-8 and Leu-11 of the I-CreI subunit to Thr and Ala, respectively, or by changing Thr-10 and Ala-13 of the I-MsoI subunit to Glu and Leu, respectively.

[0103] Techniques such as computational protein design algorithms can also be used to rationally-design the subunit interfaces. Such methods are known in the art. For example, Chevalier et al. used a computational algorithm to redesign the interface between I-CreI and the N-terminal LAGLIDADG domain of I-DmoI to enable the two to interact (Chevalier et al. (2002), Mol. Cell 10:895-905). Taking these results into account, a single-chain meganuclease comprising an N-terminal subunit derived from I-CreI and a C-terminal subunit derived from the N-terminal LAGALIDADG domain of I-DmoI can be produced by (1) selecting an N-terminal fusion point in I-CreI from Table 2, (2) selecting a C-terminal fusion point in I-DmoI from Table 10, (3) selecting a linker from Table 6 (or designing a similar linker based on the rules provided), and (4) incorporating the mutations L11A, F16I, K96N, and L97F into the I-CreI subunit and the mutations I19W, H51F, and L55R into the I-DmoI subunit as proposed by Chevalier et al.

15

[0104] Alternatively, empirical methods such as directed evolution can be used to engineer the interface between two different LAGLIDADG subunits. Such methods are known in the art. For example, genetic libraries can be produced in which specific amino acids in the subunit interface are randomized, and library members which permit the interaction between the two subunits are screened experimentally. Such screening methods are known in the art (*e.g.,* Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Chames et al. (2005), Nucl. Acids Res. 33: e178; Seligman et al. (2002), Nucl. Acids Res. 30: 3870-9, Arnould et al. (2006), J. Mol. Biol. 355: 443-58) and can be conducted to test for the ability of a single-chain meganuclease comprising two different LAGLIDADG subunits to cleave a hybrid DNA site within a yeast or bacterial cell.

6. Single-chain meganucleases with altered DNA-cleavage specificity, activity, and/or DNA-binding affinity

[0105] The invention can be used to produce single-chain meganucleases comprising individual LAGLIDADG subunits that have been genetically-engineered with respect to DNA-cleavage specificity using a variety of methods. Such methods include rational-design (*e.g.,* WO 2007/047859), computational design (*e.g.,* Ashworth et al. (2006), Nature 441:656-659), and genetic selection (Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Chames et al. (2005), Nucl. Acids Res. 33: e178; Seligman et al. (2002), Nucl. Acids Res. 30: 3870-9, Arnould et al. (2006), J. Mol. Biol. 355: 443-58). Such meganucleases can be targeted to DNA sites that differ from the sites recognized by wild-type meganucleases. The invention can also be used to join LAGLIDADG subunits that have been rationally-designed to have altered activity (*e.g.,* WO 2007/047859; Arnould et al. (2007), J. Mol. Biol 371(1):49-65) or DNA-binding affinity as described in WO 2007/047859.

7. Methods of Producing Recombinant Cells and Organisms

[0106] Aspects of the present invention further provide methods for producing recombinant, transgenic or otherwise genetically-modified cells and organisms using single-chain meganucleases. Thus, in certain embodiments, recombinant single-chain meganucleases are developed to specifically cause a double-stranded break at a single site or at relatively few sites in the genomic DNA of a cell or an organism to allow for precise insertion(s) of a sequence of interest by homologous recombination. In other embodiments, recombinant meganucleases are developed to specifically cause a double-stranded break at a single site or at relatively few sites in the genomic DNA of a cell or an organism to either (a) allow for rare insertion(s) of a sequence of interest by non-homologous end-joining or (b) allow for the disruption of the target sequence by non-homologous end-joining. As used herein with respect to homologous recombination or non-homologous end-joining of sequences of interest, the term "insertion" means the ligation of a sequence of interest into a chromosome such that the sequence of interest is integrated into the chromosome. In the case of homologous recombination, an inserted sequence can replace an endogenous sequence, such that the original DNA is replaced by exogenous DNA of equal length, but with an altered nucleotide sequence. Alternatively, an inserted sequence can include more or fewer bases than the sequence it replaces.

[0107] Therefore, in accordance with this aspect of the invention, the recombinant organisms include, but are not limited to, monocot plant species such as rice, wheat, corn (maize) and rye, and dicot species such as legumes (e.g., kidney beans, soybeans, lentils, peanuts, peas), alfalfa, clover, tobacco and *Arabidopsis* species. In addition, the recombinant organisms can include, but are not limited to, animals such as humans and non-human primates, horses, cows, goats, pigs, sheep, dogs, cats, guinea pigs, rats, mice, lizards, fish and insects such as *Drosophila* species. In other embodiments, the organism is a fungus such as a *Candida, Neurospora* or *Saccharomyces* species.

[0108] In some embodiments, the methods of the invention involve the introduction of a sequence of interest into a cell such as a germ cell or stem cell that can become a mature recombinant organism or allow the resultant genetically-modified organism to give rise to progeny carrying the inserted sequence of interest in its genome.

[0109] Meganuclease proteins can be delivered into cells to cleave genomic DNA, which allows for homologous recombination or non-homologous end-joining at the cleavage site with a sequence of interest, by a variety of different mechanisms known in the art. For example, the recombinant meganuclease protein can introduced into a cell by techniques including, but not limited to, microinjection or liposome transfections (see, *e.g.,* Lipofectamine™, Invitrogen Corp., Carlsbad, CA). The liposome formulation can be used to facilitate lipid bilayer fusion with a target cell, thereby allowing the contents of the liposome or proteins associated with its surface to be brought into the cell. Alternatively, the enzyme can be fused to an appropriate uptake peptide such as that from the HIV TAT protein to direct cellular uptake (see, *e.g.,* Hudecz et al. (2005), Med. Res. Rev. 25: 679-736).

[0110] Alternatively, gene sequences encoding the meganuclease protein are inserted into a vector and transfected into a eukaryotic cell using techniques known in the art (see, *e.g.,* Ausubel et. al., Current Protocols in Molecular Biology, Wiley 1999). The sequence of interest can be introduced in the same vector, a different vector, or by other means known in the art.

[0111] Non-limiting examples of vectors for DNA transfection include virus vectors, plasmids, cosmids, and YAC vectors. Transfection of DNA sequences can be accomplished by a variety of methods known to those of skill in the art. For instance, liposomes and immunoliposomes are used to deliver DNA sequences to cells (see, *e.g.,* Lasic et al. (1995), Science 267: 1275-76). In addition, viruses can be utilized to introduce vectors into cells (see, *e.g.,* U.S. Pat. No. 7,037,492). Alternatively, transfection strategies can be utilized such that the vectors are introduced as naked DNA (see, *e.g.,* Rui et al. (2002), Life Sci. 71(15): 1771-8).

16

[0112] General methods for delivering nucleic acids into cells include: (1) chemical methods (Graham et al. (1973), Virology 54(2):536-539; Zatloukal et al. (1992), Ann. N.Y. Acad. Sci., 660:136-153; (2) physical methods such as microinjection (Capecchi (1980), Cell 22(2):479-488, electroporation (Wong et al. (1982), Biochim. Biophys. Res. Commun. 107(2):584-587; Fromm et al. (1985), Proc. Nat'l Acad. Sci. USA 82(17):5824-5828; U.S. Pat. No. 5,384,253) and ballistic injection (Johnston et al. (1994), Methods Cell. Biol. 43(A): 353-365; Fynan et al. (1993), Proc. Nat'l Acad. Sci. USA 90(24): 11478-11482); (3) viral vectors (Clapp (1993), Clin. Perinatol. 20(1): 155-168; Lu et al. (1993), J. Exp. Med. 178(6):2089-2096; Eglitis et al. (1988), Avd. Exp. Med. Biol. 241:19-27; Eglitis et al. (1988), Biotechniques 6(7):608-614); and (4) receptor-mediated mechanisms (Curiel et al. (1991), Proc. Nat'l Acad. Sci. USA 88(19):8850-8854; Curiel et al. (1992), Hum. Gen. Ther. 3(2):147-154; Wagner et al. (1992), Proc. Nat'l Acad. Sci. USA 89 (13):6099-6103).

[0113] In certain embodiments, a genetically-modified plant is produced, which contains the sequence of interest inserted into the genome. In certain embodiments, the genetically-modified plant is produced by transfecting the plant cell with DNA sequences corresponding to the recombinant meganuclease and the sequence of interest, which may or may not be flanked by the meganuclease recognition sequences and/or sequences substantially identical to the target sequence. In other embodiments, the genetically-modified plant is produced by transfecting the plant cell with DNA sequences corresponding to the recombinant meganuclease only, such that cleavage promotes non-homologous end-joining and disrupts the target sequence containing the recognition sequence. In such embodiments, the meganuclease sequences are under the control of regulatory sequences that allow for expression of the meganuclease in the host plant cells. These regulatory sequences include, but are not limited to, constitutive plant promoters such as the NOS promoter, chemically-inducible gene promoters such as the dexamethasone-inducible promoter (see, *e.g.,* Gremillon et al. (2004), Plant J. 37:218-228), and plant tissue specific promoters such as the LGC1 promoter (see, *e.g.,* Singh et al. (2003), FEBS Lett. 542:47-52).

[0114] Suitable methods for introducing DNA into plant cells include virtually any method by which DNA can be introduced into a cell, including but not limited to *Agrobacterium* infection, PEG-mediated transformation of protoplasts (Omirulleh et al. (1993), Plant Molecular Biology, 21:415-428), desiccation/inhibition-mediated DNA uptake, electroporation, agitation with silicon carbide fibers, ballistic injection or microprojectile bombardment, and the like.

[0115] In other embodiments, a genetically-modified animal is produced using a recombinant meganuclease. As with plant cells, the nucleic acid sequences can be introduced into a germ cell or a cell that will eventually become a transgenic organism. In some embodiments, the cell is a fertilized egg, and exogenous DNA molecules can be injected into the pro-nucleus of the fertilized egg. The micro-injected eggs are then transferred into the oviducts of pseudopregnant foster mothers and allowed to develop. The recombinant meganuclease is expressed in the fertilized egg (*e.g.,* under the control of a constitutive promoter, such as 3-phosphoglycerate kinase), and facilitates homologous recombination of the sequence of interest into one or a few discrete sites in the genome. Alternatively, the genetically-modified animals can be obtained by utilizing recombinant embryonic stem ("ES") cells for the generation of the transgenics, as described by Gossler et al. (1986), Pro. Natl. Acad. Sci. USA 83:9065 9069.

[0116] In certain embodiments, a recombinant mammalian expression vector is capable of directing tissue-specific expression of the nucleic acid preferentially in a particular cell type. Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987), Genes Dev. 1: 268-277), lymphoid-specific promoters (Calame and Eaton (1988), Adv. Immunol. 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989), EMBO J. 8: 729-733) and immunoglobulins (Banerji et al. (1983), Cell 33: 729-740; Queen and Baltimore (1983), Cell 33: 741-748), neuron-specific promoters (*e.g.,* the neurofilament promoter; Byrne and Ruddle (1989), Proc. Natl. Acad. Sci. USA 86: 5473-5477), pancreas-specific promoters (Edlund et al. (1985), Science 230: 912-916), and mammary gland-specific promoters (*e.g.,* milk whey promoter; U.S. Pat. No. 4,873,316 and European Pat. Pub. EP 0 264 166). Developmentally-regulated promoters are also encompassed, *e.g.,* the murine hox promoters (Kessel and Gruss (1990), Science 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989), Genes Dev. 3: 537-546).

[0117] In certain embodiments, a single-chain meganuclease may be tagged with a peptide epitope (*e.g.,* an HA, FLAG, or Myc epitope) to monitor expression levels or localization. In some embodiments, the meganuclease may be fused to a sub-cellular localization signal such as a nuclear-localization signal (*e.g.,* the nuclear localization signal from SV40) or chloroplast or mitochondrial localization signals. In other embodiments, the meganuclease may be fused to a nuclear export signal to localize it to the cytoplasm. The meganuclease may also be fused to an unrelated protein or protein domain such as a protein that stimulates DNA-repair or homologous recombination (e.g., recA, RAD51, RAD52, RAD54, RAD57 or BRCA2).

8. Methods for Gene Therapy

[0118] Aspects of the invention allow for the use of recombinant meganuclease for gene therapy. As used herein, "gene therapy" means therapeutic treatments that comprise introducing into a patient a functional copy of at least one gene, or gene regulatory sequence such as a promoter, enhancer, or silencer to replace a gene or gene regulatory region that is defective in its structure and/or function. The term "gene therapy" can also refer to modifications made to a deleterious gene or regulatory element (e.g., oncogenes) that reduce or eliminate expression of the gene. Gene therapy can be performed to treat congenital conditions, conditions resulting from mutations or damage to specific genetic loci over the life of the patient, or conditions resulting from infectious organisms.

[0119] In some aspects of the invention, dysfunctional genes are replaced or disabled by the insertion of exogenous nucleic acid sequences into a region of the genome affecting gene expression. In certain embodiments, the recombinant meganuclease is targeted to a particular sequence in the region of the genome to be modified so as to alleviate the condition. The sequence can be a region within an exon, intron, promoter, or other regulatory region that is causing dysfunctional expression of the gene. As used herein, the term "dysfunctional expression" means aberrant expression of a gene product either by the cell producing too little of the gene product, too much of the gene product, or producing a gene product that has a different function such as lacking the necessary function or having more than the necessary function.

[0120] Exogenous nucleic acid sequences inserted into the modified region can be used to provide "repaired" sequences that normalize the gene. Gene repair can be accomplished by the introduction of proper gene sequences into the gene allowing for proper function to be reestablished. In these embodiments, the nucleic acid sequence to be inserted can be the entire coding sequence for a protein or, in certain embodiments, a fragment of the gene comprising only the region to be repaired. In other embodiments the nucleic acid sequence to be inserted comprises a promoter sequence or other regulatory elements such that mutations causing abnormal expression or regulation are repaired. In other embodiments, the nucleic acid sequence to be inserted contains the appropriate translation stop codon lacking in a mutated gene. The nucleic acid sequence can also have sequences for stopping transcription in a recombinant gene lacking appropriate transcriptional stop signals.

[0121] Alternatively, the nucleic acid sequences can eliminate gene function altogether by disrupting the regulatory sequence of the gene or providing a silencer to eliminate gene function. In some embodiments, the exogenous nucleic acid sequence provides a translation stop codon to prevent expression of the gene product. In other embodiments, the exogenous nucleic acid sequences provide transcription stop element to prevent expression of a full length RNA molecule. In still other embodiments, gene function is disrupted directly by the meganuclease by introducing base insertions, base deletions, and/or frameshift mutations through non-homologous end-joining.

[0122] In many instances, it is desirable to direct the proper genetic sequences to a target cell or population of cells that is the cause of the disease condition. Such targeting of therapeutics prevents healthy cells from being targeted by the therapeutics. This increases the efficacy of the treatment, while decreasing the potentially adverse effects that the treatment could have on healthy cells.

[0123] Delivery of recombinant meganuclease genes and the sequence of interest to be inserted into the genome to the cells of interest can be accomplished by a variety of mechanisms. In some embodiments, the nucleic acids are delivered to the cells by way of viruses with particular viral genes inactivated to prevent reproduction of the virus. Thus, a virus can be altered so that it is capable only of delivery and maintenance within a target cell, but does not retain the ability to replicate within the target cell or tissue. One or more DNA sequences can be introduced to the altered viral genome, so as to produce a viral genome that acts like a vector, and may or may not be inserted into a host genome and subsequently expressed. More specifically, certain embodiments include employing a retroviral vector such as, but not limited to, the MFG or pLJ vectors. An MFG vector is a simplified Moloney murine leukemia virus vector (MoMLV) in which the DNA sequences encoding the pol and env proteins have been deleted to render it replication defective. A pLJ retroviral vector is also a form of the MoMLV (see, *e.g.,* Korman et al. (1987), Proc. Nat'l Acad. Sci., 84:2150-2154). In other embodiments, a recombinant adenovirus or adeno-associated virus can be used as a delivery vector.

[0124] In other embodiments, the delivery of recombinant meganuclease protein and/or recombinant meganuclease gene sequences to a target cell is accomplished by the use of liposomes. The production of liposomes containing nucleic acid and/or protein cargo is known in the art (see, *e.g.,* Lasic et al. (1995), Science 267: 1275-76). Immunoliposomes incorporate antibodies against cell-associated antigens into liposomes, and can delivery DNA sequences for the meganuclease or the meganuclease itself to specific cell types (see, *e.g.,* Lasic et al. (1995), Science 267: 1275-76; Young et al. (2005), J. Calif. Dent. Assoc. 33(12): 967-71; Pfeiffer et al. (2006), J. Valc. Surg. 43(5):1021-7). Methods for producing and using liposome formulations are well known in the art, (see, *e.g.,* U.S. Pat. No. 6,316,024, U.S. Pat. No. 6,379,699, U.S. Pat. No. 6,387,397, U.S. Pat. No. 6,511,676 and U.S. Pat. No. 6,593,308, and references cited therein). In some embodiments, liposomes are used to deliver the sequences of interest as well as the recombinant meganuclease protein or recombinant meganuclease gene sequences.

9. Methods for Treating Pathogen Infection.

[0125] Aspects of the invention also provide methods of treating infection by a pathogen. Pathogenic organisms include viruses such as, but not limited to, herpes simplex virus 1, herpes simplex virus 2, human immunodeficiency virus 1, human immunodeficiency virus 2, variola virus, polio virus, Epstein-Barr virus, and human papilloma virus and bacterial organisms such as, but not limited to, *Bacillus anthracis, Haemophilus* species, *Pneumococcus* species, *Staphylococcus aureus, Streptococcus* species, methicillin-resistant *Staphylococcus aureus,* and *Mycoplasma tuberculosis.* Pathogenic organisms also include fungal organisms such as, but not limited to, *Candida, Blastomyces, Cryptococcus,* and *Histoplasma* species.

[0126] In some embodiments, a single-chain meganuclease can be targeted to a recognition sequence within the pathogen genome, *e.g.,* to a gene or regulatory element that is essential for growth, reproduction, or toxicity of the pathogen. In certain embodiments, the recognition sequence may be in a bacterial plasmid. Meganuclease-mediated cleavage of a recognition sequence in a pathogen genome can stimulate mutation within a targeted, essential gene in the form of an insertion, deletion or frameshift, by stimulating non-homologous end-joining.

Alternatively, cleavage of a bacterial plasmid can result in loss of the plasmid along with any genes encoded on it, such as toxin genes (*e.g., B. anthracis* Lethal Factor gene) or antibiotic resistance genes. As noted above, the meganuclease may be delivered to the infected patient, animal, or plant in either protein or nucleic acid form using techniques that are common in the art. In certain embodiments, the meganuclease gene may be incorporated into a bacteriophage genome for delivery to pathogenic bacteria.

[0127] Aspects of the invention also provide therapeutics for the treatment of certain forms of cancer. Because human viruses are often associated with tumor formation (e.g., Epstein-Barr Virus and nasopharyngeal carcinomas; Human Papilloma Virus and cervical cancer) inactivation of these viral pathogens may prevent cancer development or progression. Alternatively, double-stranded breaks targeted to the genomes of these tumor-associated viruses using single-chain meganucleases may be used to trigger apoptosis through the DNA damage response pathway. In this manner, it may be possible to selectively induce apoptosis in tumor cells harboring the viral genome.

10. Methods for Genotyping and Pathogen Identification

[0128] Aspects of the invention also provide tools for *in vitro* molecular biology research and development. It is common in the art to use site-specific endonucleases (e.g., restriction enzymes) for the isolation, cloning, and manipulation of nucleic acids such as plasmids, PCR products, BAC sequences, YAC sequences, viruses, and genomic sequences from eukaryotic and prokaryotic organisms (see, *e.g.,* Ausubel et al., Current Protocols in Molecular Biology, Wiley 1999). Thus, in some embodiments, a single-chain meganuclease may be used to manipulate nucleic acid sequences *in vitro.* For example, single-chain meganucleases recognizing a pair of recognition sequences within the same DNA molecule can be used to isolate the intervening DNA segment for subsequent manipulation such as ligation into a bacterial plasmid, BAC, or YAC.

[0129] In another aspect, this invention provides tools for the identification of pathogenic genes and organisms. In one embodiment, single-chain meganucleases can be used to cleave recognition sites corresponding to polymorphic genetic regions correlated to disease to distinguish disease-causing alleles from healthy alleles *(e.g.,* a single-chain meganuclease which recognizes the ΔF-508 allele of the human CFTR gene, see example 4). In this embodiment, DNA sequences isolated from a human patient or other organism are digested with a single-chain meganuclease, possibly in conjunction with additional site-specific nucleases, and the resulting DNA fragment pattern is analyzed by gel electrophoresis, capillary electrophoresis, mass spectrometry, or other methods known in the art. This fragmentation pattern and, specifically, the presence or absence of cleavage by the single-chain meganuclease, indicates the genotype of the organism by revealing whether or not the recognition sequence is present in the genome. In another embodiment, a single-chain meganuclease is targeted to a polymorphic region in the genome of a pathogenic virus, fungus, or bacterium and used to identify the organism. In this embodiment, the single-chain meganuclease cleaves a recognition sequence that is unique to the pathogen (*e.g.,* the spacer region between the 16S and 23S rRNA genes in a bacterium; see, *e.g.,* van der Giessen et al. (1994), Microbiology 140:1103-1108) and can be used to distinguish the pathogen from other closely-related organisms following endonuclease digest of the genome and subsequent analysis of the fragmentation pattern by electrophoresis, mass spectrometry, or other methods known in the art.

11. Methods for the Production of Custom DNA-binding Domains.

[0130] In another aspect, the invention provides single-chain DNA-binding proteins that lack endonuclease cleavage activity. The catalytic activity of a single-chain meganuclease can be eliminated by mutating amino acids involved in catalysis (e.g., the mutation of Q47 to E in I-CreI, see Chevalier et al. (2001), Biochemistry. 43:14015-14026); the mutation of D44 or D 145 to N in I-SceI; the mutation of E66 to Q in I-CeuI; the mutation of D22 to N in I-MsoI). The inactivated meganuclease can then be fused to an effector domain from another protein including, but not limited to, a transcription activator *(e.g.,* the GAL4 transactivation domain or the VP16 transactivation domain), a transcription repressor (e.g., the KRAB domain from the Kruppel protein), a DNA methylase domain *(e.g.,* M.CviPI or M.SssI), or a histone acetyltransferase domain *(e.g.,* HDAC1 or HDAC2). Chimeric proteins consisting of an engineered DNA-binding domain, most notably an engineered zinc finger domain, and an effector domain are known in the art (see, *e.g.,* Papworth et al. (2006), Gene 366:27-38).

EXAMPLES

[0131] This invention is further illustrated by the following examples, which should not be construed as limiting. Example 1 presents evidence that a previously disclosed method for the production of single-chain I-CreI meganucleases (Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62 ; WO 2003/078619) is not sufficient for the production of meganucleases recognizing non-palindromic DNA sites. Examples 2 and 3 present evidence that the method described here is sufficient to produce single-chain I-CreI meganucleases recognizing non-palindromic DNA sites using a flexible Gly-Ser linker (example 2) or a designed, structured linker (example 3). Although examples 2 and 3 below refer specifically to single-chain meganucleases based on I-CreI, single-chain meganucleases comprised of subunits derived from I-SceI, I-MsoI, I-CeuI, I-AniI, and other LAGLIDADG meganucleases can be similarly produced and used, as described herein. Examples not falling within the scope of the claims are for illustrative purposes only.

EXAMPLE 1

Evaluation of the Method of *Epinat et al.*

1. Single chain meganucleases using the method of Epinat *et al.*

[0132] Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62 and WO 2003/078619 report the production of a single-chain meganuclease derived from the I-CreI meganuclease. Specifically, the authors used an 11 amino acid peptide linker derived from I-DmoI (amino acids 94-104 of I-DmoI, sequence MLERIRLFNMR) to join an N-terminal I-CreI subunit (amino acids 1-93 of I-CreI) to a C-terminal I-CreI subunit (amino acids 8-163). This particular arrangement of N-terminal subunit - linker - C-terminal subunit was selected because it most closely mimics the domain organization of the di-LAGLIDADG I-DmoI meganuclease. The authors evaluated the single-chain I-CreI meganuclease experimentally and found it to cleave a wild-type I-CreI recognition sequence effectively, albeit at a significantly reduced rate relative to the wild-type I-CreI homodimer.

[0133] Because the fusion protein produced by these authors comprised two otherwise wild-type subunits, both of which recognize identical DNA half-sites, it was necessary to test the single-chain meganuclease using the pseudo-palindromic wild-type DNA site. As such, it was not possible for the authors to rule out the possibility that the observed cleavage activity was not due to cleavage by an individual single-chain meganuclease but, rather, by a intermolecular dimer of two single-chain meganucleases in which one domain from each associated to form a functional meganuclease that effectively behaves like the wild-type homodimer. Indeed, a substantial portion of the N-terminal I-CreI subunit (amino acids 94-163) was removed in the production of the single-chain meganuclease reported by *Epinat et al.* An inspection of the three-dimensional I-CreI crystal structure (Jurica et al. (1998), Mol. Cell 2:469-476) reveals that this truncation results in the removal of three alpha-helices from the surface of the N-terminal subunit and the subsequent exposure to solvent of a significant amount of hydrophobic surface area. As such, the present inventors hypothesized that the N-terminal subunit from the single-chain I-CreI meganuclease of *Epinat et al.* is unstable and inactive and that the observed DNA cleavage activity is, in fact, due to the dimerization of the C-terminal subunits from two single-chain proteins. The protein stability problems resulting from application of the method *of Epinat et al.* are also discussed in Fajardo-Sanchez et al. (2008), Nucleic Acids Res. 36:2163-2173.

2. Design of single-chain LAM meganucleases using the method of Epinat *et al.*

[0134] To more critically evaluate the method for single-chain I-CreI meganuclease production reported by Epinat *et al. (*Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62; WO 2003/078619), a single-chain meganuclease was produced in which the N- and C-terminal I-CreI domains recognize different DNA half-sites. The method reported in Epinat *et al.* was used to produce a pair of single-chain meganucleases comprising one LAM1 domain and one LAM2 domain. This "LAM1epLAM2" meganuclease (SEQ ID NO: 48) comprises an N-terminal LAM1 domain and a C-terminal LAM2 domain while "LAM2epLAM1" (SEQ ID NO: 49) comprises an N-terminal LAM2 domain and a C-terminal LAM1 domain. In total, both single-chain meganucleases differ by 11 amino acids from that reported by *Epinat et al.* and all amino acid changes are in regions of the enzyme responsible for DNA recognition which are not expected to affect subunit interaction.

3. Construction of single-chain Meganucleases.

[0135] LAM1epLAM2 and LAM2epLAM1 were produced by PCR of existing LAM1 and LAM2 genes with primers that introduce the I-DmoI linker sequence (which translates to MLERIRLFNMR) as well as restriction enzyme sites for cloning. The two LAM subunits were cloned sequentially into pET-21a vectors with a six histidine tag fused at the 3' end of the full-length single-chain gene for purification (Novagen Corp., San Diego, CA). All nucleic acid sequences were confirmed using Sanger Dideoxynucleotide sequencing (see, Sanger et al. (1977), Proc. Natl. Acad. Sci. USA. 74(12): 5463-7).

[0136] The LAMep meganucleases were expressed and purified using the following method. The constructs cloned into a pET21a vector were transformed into chemically competent BL21 (DE3) pLysS, and plated on standard 2xYT plates containing 200 µg/ml carbanicillin. Following overnight growth, transformed bacterial colonies were scraped from the plates and used to inoculate 50 ml of 2XYT broth. Cells were grown at 37°C with shaking until they reached an optical density of 0.9 at a wavelength of 600 nm. The growth temperature was then reduced from 37°C to 22°C. Protein expression was induced by the addition of 1 mM IPTG, and the cells were incubated with agitation for two and a half hours. Cells were then pelleted by centrifugation for 10 min. at 6000x g. Pellets were resuspended in 1 ml binding buffer (20 mM Tris-HCL, pH 8.0, 500 mM NaCl, 10 mM imidazole) by vortexing. The cells were then disrupted with 12 pulses of sonication at 50% power and the cell debris was pelleted by centrifugation for 15 min at 14,000x g. Cell supernatants were diluted in 4 ml binding buffer and loaded onto a 200 µl nickel-charged metal-chelating Sepharose column (Pharmacia).

[0137] The column was subsequently washed with 4 ml wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 60 mM imidazole) and with 0.2 ml elution buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 400 mM imidazole). Meganuclease enzymes were eluted with an additional 0.6 ml of

elution buffer and concentrated to 50-130 µl using Vivospin disposable concentrators (ISC, Inc., Kaysville, UT). The enzymes were exchanged into SA buffer (25 mM Tris-HCL, pH 8.0, 100 mM NaCl, 5 mM MgCl$_2$, 5 mM EDTA) for assays and storage using Zeba spin desalting columns (Pierce Biotechnology, Inc., Rockford, IL). The enzyme concentration was determined by absorbance at 280 nm using an extinction coefficient of 23,590 M^{-1}cm^{-1}. Purity and molecular weight of the enzymes was then confirmed by MALDI-TOF mass spectrometry.

4. Cleavage Assays.

[0138] All enzymes purified as described above were assayed for activity by incubation with linear, double-stranded DNA substrates containing meganuclease recognition sequences. Synthetic oligonucleotides corresponding to both sense and antisense strands of the recognition sequences were annealed and were cloned into the *SmaI* site of the pUC19 plasmid by blunt-end ligation. The sequences of the cloned binding sites were confirmed by Sanger dideoxynucleotide sequencing. All plasmid substrates were linearized with *XmnI* or *ScaI* concurrently with the meganuclease digest. The enzyme digests contained 5 µl 0.05 µM DNA substrate, 2.5 µl 5 µM single-chain meganuclease, 9.5 µl SA buffer, and 0.5 µl *XmnI* or *ScaI*. Digests were incubated at either 37°C for four hours. Digests were stopped by adding 0.3 mg/ml Proteinase K and 0.5% SDS, and incubated for one hour at 37°C. Digests were analyzed on 1.5% agarose and visualized by ethidium bromide staining.

5. Results

[0139] The LAMep meganucleases produced using the method of Epinat *et al.* were incubated with DNA substrates comprising the LAM1 palindrome (SEQ ID NOs: 40 and 41), the LAM2 palindrome (SEQ ID NOs. 44 and 45), or the LAM1/LAM2 hybrid site (SEQ ID NOs. 46 and 47). The LAM1epLAM2 single-chain meganuclease was found to cleave primarily the LAM2 palindrome whereas the LAM2epLAM1 single-chain meganuclease was found to cleave primarily the LAM1 palindrome. Neither single-chain meganuclease cleaved the hybrid site to a significant degree. These results suggest that, indeed, the method of Epinat *et al.* produces single-chain meganucleases that are unable to cleave non-palindromic DNA sequences. Both single-chain meganucleases were found to cleave primarily the recognition sequence corresponding to a palindrome of the half-site recognized by the C-terminal subunit, suggesting that the N-terminal subunit is inactive. Thus, the active meganuclease species characterized by Epinat *et al.* appears to be primarily a dimer between the C-terminal subunits of a pair of single-chain I-CreI meganucleases. Alternatively, cleavage of the palindromic DNA site may be due to sequential single strand nicking by the C-terminal subunits of different single-chain I-CreI meganucleases. In either case, in contrast to claims made by *Epinat et al.,* the method does not produce a substantially functional single-chain I-CreI heterodimer and is generally not useful for the recognition and cleavage of non-palindromic DNA sites.

EXAMPLE 2

Single-Chain I-CreI Meganucleases Produced Using a Flexible Gly-Ser Linker

1. Design of single-chain LAM meganucleases using a Gly-Ser linker

[0140] The designed LAM1 and LAM2 endonucleases were fused into a single polypeptide using Linker 3 from Table 3. Val-151 was used as the N-terminal fusion point (to the LAM1 subunit) while Phe-9 was the C-terminal fusion point (to the LAM2 subunit). The resulting single-chain meganuclease, "LAM1gsLAM2" (SEQ ID NO: 50) was cloned into pET21 a, expressed in *E. coli* and purified as described in Example 1.

2. Results

[0141] LAM1gsLAM2 was assayed for cleavage activity using the same DNA substrates and incubation conditions as described in Example 1. In contrast to results with the LAMep meganucleases, LAM1gsLAM2 was found to cleave primarily the hybrid LAM1/LAM2 recognition sequence (SEQ ID NOs: 46 and 47). The extent of cleavage is significantly reduced relative to the LAM1/LAM2 heterodimer produced by co-expressing the LAM1 and LAM2 monomers in *E. coli.* Under the same reaction conditions, the heterodimer cleaves the LAM1/LAM2 recognition sequence to completion, suggesting that the Gly-Ser linker impairs cleavage activity to some extent. Nonetheless, LAM1gsLAM2 exhibits a much stronger preference for the hybrid site over the palindromic LAM1 or LAM2 sites and, so has utility for applications in which specificity is of greater importance than activity.

EXAMPLE 3

Single-Chain I-Crel Meganucleases Produced Using a Structured Linker

1. Design of single-chain LAM meganucleases using a designed, structured linker

[0142] The designed LAM1 and LAM2 endonucleases were fused into a single polypeptide using Linker 9 from Table 6. Asp-153 was used as the N-terminal fusion point (to the LAM1 subunit) while Lys-7 was the C-terminal fusion point (to the LAM2 subunit). The resulting single-chain meganuclease, "LAM1desLAM2" (SEQ ID NO: 51) was cloned into pET21 a, expressed in *E. coli* and purified as described in Example 1.

2. Results

[0143] LAM1desLAM2 was assayed for cleavage activity using the same DNA substrates and incubation conditions as described in Example 1. In contrast to results with the LAMep meganucleases, LAM1desLAM2 was found to cleave primarily the hybrid LAM1/LAM2 recognition sequence (SEQ ID NO: 46 and 47). The extent of cleavage is comparable to the LAM1/LAM2 heterodimer produced by co-expressing the LAM1 and LAM2 monomers in *E. coli*. These results suggest that designed, structured linkers such as Linker 9 do not interfere significantly with cleavage activity. Moreover, LAM1desLAM2 is structurally stable and maintains catalytic activity for >3 weeks when stored in SA buffer at 4°C. Importantly, LAM1desLAM2 exhibits minimal activity toward the palindromic LAM1 and LAM2 sites (SEQ ID NOS: 40 and 41 and 44 and 45), indicating that the functional species produced by the method disclosed here is primarily a single-chain heterodimer.

EXAMPLE 4

Single-Chain I-Msol Meganucleases Produced Using a Structured Linker

1. Design of single-chain I-Msol meganucleases using a designed, structured linker

[0144] A pair of I-Msol endonuclease subunits (unmodified with respect to DNA cleavage specificity) were fused into a single polypeptide using Linker 30 from Table 8. Ile-166 was used as the N-terminal fusion point while Leu-7 was the C-terminal fusion point. The resulting single-chain meganuclease, "MSOdesMSO" (SEQ ID NO: 52) was cloned into pET21a with a C-terminal 6xHis-tag to facilitate purification. The meganuclease was then expressed in *E. coli* and purified as described in Example 1.

2. Results

[0145] Purified MSOdesMSO was assayed for the ability to cleave a plasmid substrate harboring the wild-type I-Msol recognition sequence (SEQ ID NO:53 and SEQ ID NO:54 and 54) under the incubation conditions as described in Example 1. The enzyme was found to have cleavage activity comparable to the I-Msol homodimer (which, in this case, is expected to recognize and cut the same recognition sequence as MSOdesMSO). SDS-PAGE analyses revealed that MSOdesMSO has an apparent molecular weight of ~40 kilodaltons, consistent with it being a pair of covalently joined I-Msol subunits, and no protein degradation products were apparent. These results indicate that the invention is suitable for the production of stable, high-activity single-chain meganucleases derived from I-Msol.

TABLE 11

I-Crel Modifications from WO 2007/047859										
	Favored Sense-Strand Base									
Posn.	A	C	G	T	A/T	A/C	A/G	C/T	G/T	A/G/T	A/C/G/T
-1	Y75	**R70***	K70	Q70*				**T46***			G70
	L75*	H75*	E70*	C70							A70
	C75*	R75*	E75*	L70							S70
	Y139*	H46*	E46*	Y75*							G46*
	C46*	K46*	D46*	Q75*							
	A46*	R46*		H75*							
				H139							

	I-CreI Modifications from WO 2007/047859										
	Favored Sense-Strand Base										
Posn.	A	C	G	T	A/T	A/C	A/G	C/T	G/T	A/G/T	A/C/G/T
				Q46*							
				H46*							
-2	Q70	E70	H70	**Q44***	C44*						
	T44*	D70	D44*								
	A44*	K44*	E44*								
	V44*	R44*									
	I44*										
	L44*										
	N44*										
-3	Q68	E68	R68	M68		H68		Y68	K68		
	C24*	F68		C68							
	I24*	K24*		L68							
		R24*		F68							
-4	A26*	E77	R77					S77			S26*
	Q77	K26*	E26*					**Q26***			
-5		E42	R42			**K28***	C28*				M66
							Q42				K66
-6	Q40	E40	R40	C40	A40						**S40**
	C28*	R28*		I40	A79						S28*
				V40	A28*						
				C79	H28*						
				I79							
				V79							
				Q28*							
-7	**N30***	E38	K38	I38			C38				H38
	Q38	K30*	R38	L38							N38
		R30*	E30*								Q30*
-8	F33	E33	F33	L33		R32*	R33				
	Y33	D33	H33	V33							
				I33							
				F33							
				C33							
-9		E32	R32	L32				D32			**S32**
			K32	V32				I32			N32
				A32							H32
				C32							Q32
											T32

Bold entries are wild-type contact residues and do not constitute "modifications" as used herein.

[0146] An asterisk indicates that the residue contacts the base on the antisense strand.
TABLE 12

	I-Msol Modifications from WO 2007/047859			
	Favored Sense-Strand Base			
Position	A	C	G	T
-1	K75*	D77	K77	C77

I-MsoI Modifications from WO 2007/047859			
Favored Sense-Strand Base			
Q77	E77	R77	L77
A49*	K49*	E49*	Q79*
C49*	**R75***	E79*	
K79*	K75*		
	R79*		
	K79*		

-2	Q75	E75	K75	A75
	K81	D75	E47*	C75
	C47*	R47*	E81*	V75
	I47*	K47*		I75
	L47*	K81*		T75
		R81*		Q47*
				Q81*
-3	Q72	E72	**R72**	K72
	C26*	Y72	K72	Y72
	L26*	H26*	**Y26***	H26*
	V26*	K26*	F26*	
	A26*	R26*		
	I26*			
-4	**K28**	**K28***	R83	**K28**
	Q83	R28*	K83	K83
		E83		Q28*
-5	**K28**	**K28***	R45	Q28*
	C28*	R28*	E28*	
	L28*			
	I28*			
-6	**I30***	E43	R43	K43
	V30*	E85	K43	**185**
	S30*	K30*	K85	V85
	L30*	R30*	R85	L85
	Q43		E30*	Q30*
			D30*	
-7	**Q41**	E32	**R32**	K32
		E41	R41	M41
			K41	L41
				141
-8	**Y35**	E32	**R32**	K32
	K35		K32	K35
			K35	
			R35	
-9	N34	**D34**	K34	S34
	H34	E34	R34	C34
		S34	H34	V34
				T34
				A34

Bold entries are represent wild-type contact residues and do not constitute "modifications" as used herein. An asterisk indicates that the residue contacts the base on the antisense strand.

TABLE 13

I-Ceu Modifications from WO 2007/047859			
	Favored Sense-Strand Base		
Position	A	C	G	T
-1	C92*	**K116***	E116*	Q116*
	A92*	R116*	E92*	Q92*
	V92*	D116*		
		K92*		
-2	Q117	E117	K117	C117
	C90*	D117	R124	V117
	L90*	R174*	K124	T117
	V90*	K124*	E124*	Q90*
		K90*	E90*	
		R90*	D90*	
		K68*		
-3	C70*	K70*	E70*	Q70*
	V70*		**E88***	
	T70*			
	L70*			
	K70*			
-4	Q126	E126	R126	K126
	N126	D126	K126	L126
	K88*	R88*	**E88***	Q88*
	L88*	K88*	D88*	
	C88*	K72*		
	C72*			
	L72*			
	V72*			
-5	C74*	**K74***	E74*	C128
	L74*		**K128**	L128
	V74*		R128	V128
	T74*		E128	T128
-6	Q86	**D86**	**K128**	K86
		E86	R128	C86
		R84*	R86	L86
		K84*	K86	
			E84*	
-7	**L76***	R76*	E76*	H76*
	C76*	K76*	R84	Q76*
	K76*	H76*		
-8	Y79	D79	R79	C79
	R79	E79	K79	L79
	Q76	D76	K76	V79
		E76	R76	L76
-9	Q78	D78	R78	K78
	N78	E78	K78	V78

25

I-Ceu Modifications from WO 2007/047859			
	Favored Sense-Strand Base		
Position	**A**	**C**	**G**	**T**

Position	**A**	**C**	**G**	**T**
	H78		H78	L78
	K78			C78
				T78

Bold entries are wild-type contact residues and do not constitute "modifications" as used herein.

[0147] An asterisk indicates that the residue contacts the base on the antisense strand.

TABLE 14

I-Scel Modifications from WO 2007/047859			
	Favored Sense-Strand Base		

Position	**A**	**C**	**G**	**T**
4	K50	**R50***	E50*	K57
		K50*	R57	M57
		E57	K57	Q50*
5	K48	**R48***	E48*	Q48*
	Q102	K48*	K102	C102
		E102	R102	L102
		E59		V102
6	K59	R59*	K84	**Q59***
		K59*	E59*	Y46
7	C46*	R46*	**K86**	K68
	L46*	K46*	R86	C86
	V46*	E86	E46*	L86 Q46*
8	K61*	E88	**E61***	K88
	S61*	R61*	**R88**	Q61*
	V61*	H61*	K88	H61*
	A61*			
	L61*			
9	**T98***	R98*	E98*	Q98*
	C98*	K98*	D98*	
	V98*			
	L98*			
10	**V96***	K96*	D96*	Q96*
	C96*	R96*	E96*	
	A96*			
11	C90*	K90*	E90*	Q90*
	L90*	R90*		
12	Q193	E165	K165	C165
		E193	R165	L165
		D193		C193
				V193
				A193
				T193
				S193
13	C193*	**K193***	E193*	Q193*
	L193*	R193*	D193*	C163

I-SceI Modifications from WO 2007/047859				
	Favored Sense-Strand Base			
Position	A	C	G	T
		D192	K163	L163
			R192	
14	L192*	E161	K147	K161
	C192*	R192*	K161	Q192*
		K192*	R161	
			R197	
			D192*	
			E192*	
15		E151	K151	C151
				L151
				K151
17	**N152***	K152*	N152*	Q152*
	S152*	K150*	S152*	Q150*
	C150*		D152*	
	L150*		**D150***	
	V150*		E150*	
	T150*			
18	K155*	R155*	E155*	H155*
	C155*	K155*		Y155*
Bold entries are wild-type contact residues and do not constitute "modifications" as used herein.				

[0148] An asterisk indicates that the residue contacts the base on the antisense strand.

[0149] Disclosed are the following items:

1. 1. A recombinant single-chain meganuclease comprising:

   a first LAGLIDADG subunit derived from a first mono-LAGLIDADG meganuclease, said first LAGLIDADG subunit having a first recognition half-site;

   a second LAGLIDADG subunit derived from a second mono-LAGLIDADG meganuclease or a di-LAGLIDADG meganuclease, said second LAGLIDADG subunit having a second recognition half-site;

   wherein said first and second LAGLIDADG subunits are covalently joined by a polypeptide linker such that said first LAGLIDADG domain is N-terminal to said linker and said second LAGLIDADG domain is C-terminal to said linker; and

   wherein said first and second LAGLIDADG subunits are capable of functioning together to recognize and cleave a non-palindromic DNA sequence which is a hybrid of said first recognition half-site and said second recognition half-site.

2. 2. The recombinant single-chain meganuclease of item 1 wherein:

   the first LAGLIDADG subunit is derived from a mono-LAGLIDADG meganuclease selected from the group consisting of I-CreI, I-MsoI and I-CeuI; and

   the second LAGLIDADG subunit is derived from either (1) a mono-LAGLIDADG meganuclease selected from the group consisting of I-CreI, I-MsoI and I-CeuI, or (2) a di-LAGLIDADG meganuclease selected from the group consisting of I-DmoI, I-SceI and I-AniI.

3. 3. The recombinant single-chain meganuclease of item 1 wherein:

   the first LAGLIDADG subunit is derived from a different species than the second LAGLIDADG subunit.

4. 4. The recombinant single-chain meganuclease of item 1 wherein:

   said first LAGLIDADG subunit comprises a polypeptide sequence having at least 85% sequence identity to a first LAGLIDADG domain selected from the group consisting of residues 9-151 of a wild-type I-CreI meganuclease of SEQ ID NO: 1; residues 11-162 of a wild-type I-MsoI meganuclease of SEQ ID NO: 2; and residues 55-210 of a wild-type I-CeuI meganuclease of SEQ ID NO: 3.

5. 5. The recombinant single-chain meganuclease of item 2 wherein:

said second LAGLIDADG subunit comprises a polypeptide sequence having at least 85% sequence identity to a second LAGLIDADG domain selected from the group consisting of residues 9-151 of a wild-type I-CreI meganuclease of SEQ ID NO: 1; residues 11-162 of a wild-type I-MsoI meganuclease of SEQ ID NO: 2; residues 55-210 of a wild-type I-CeuI meganuclease of SEQ ID NO: 3; residues 9-96 of a wild-type I-DmoI of SEQ ID NO: 4; residues 105-178 of a wild-type I-DmoI of SEQ ID NO: 4; residues 32-123 of a wild-type I-SceI of SEQ ID NO: 5; residues 134-225 of a wild-type I-SceI of SEQ ID NO: 5; residues 4-121 of a wild-type I-AniI of SEQ ID NO: 6; and residues 136-254 of a wild-type I-AniI of SEQ ID NO: 6.

6. 6. The recombinant single-chain meganuclease of item 2 wherein:

each of said LAGLIDADG subunits comprises at least 85% identity to a LAGLIDADG domain independently selected from the group consisting of residues 9-151 of a wild-type I-CreI meganuclease of SEQ ID NO: 1; residues 11-162 of a wild-type I-MsoI meganuclease of SEQ ID NO: 2; residues 55-210 of a wild-type I-CeuI meganuclease of SEQ ID NO: 3; residues 9-96 of a wild-type I-DmoI of SEQ ID NO: 4; residues 105-178 of a wild-type I-DmoI of SEQ ID NO: 4; residues 32-123 of a wild-type I-SceI of SEQ ID NO: 5; residues 134-225 of a wild-type I-SceI of SEQ ID NO: 5; residues 4-121 of a wild-type I-AniI of SEQ ID NO: 6; and residues 136-254 of a wild-type I-AniI of SEQ ID NO: 6; and

at least one of said LAGLIDADG domains comprises at least one amino acid modification disclosed in any of Tables 11, 12, 13 and 14.

7. 7. The recombinant single-chain meganuclease of item 6 wherein:

at least one LAGLIDADG domain is derived from I-CreI and at least one modification is selected from Table 1 of any of Tables 11, 12, 13 and 14;

at least one LAGLIDADG domain is derived from I-MsoI and at least one modification is selected from Table 12;

at least one LAGLIDADG domain is derived from I-CeuI and at least one modification is selected from Table 13; or

at least one LAGLIDADG domain is derived from I-SceI and at least one modification is selected from Table 14.

8. 8. The recombinant single-chain meganuclease of item 2 wherein:

each of said LAGLIDADG subunits has a recognition half-site selected from the group consisting of SEQ ID NOs: 7-30.

9. 9. The recombinant single-chain meganuclease of item 8 wherein:

at least one of said LAGLIDADG subunits has a recognition half-site selected from the group consisting of SEQ ID NOs: 7-30; and

the other of said LAGLIDADG subunits has a recognition half-site which differs by at modification of at least one base pair from a recognition half-site selected from the group consisting of SEQ ID NOs: 7-30.

10. 10. The recombinant single-chain meganuclease of any one of items 1-9 wherein:

said polypeptide linker is a flexible linker.

11. 11. The recombinant single-chain meganuclease of item 10 wherein:

said linker comprises 15-40 residues.

12. 12. The recombinant single-chain meganuclease of item 10 wherein:

said linker comprises 25-31 residues.

13. 13. The recombinant single-chain meganuclease of item 10 wherein:

at least 50% of said linker comprises polar uncharged residues.

14. 14. The recombinant single-chain meganuclease of any one of items 1-9 wherein:

said polypeptide linker has a stable secondary structure.

15. 15. The recombinant single-chain meganuclease of item 14 wherein:

said stable secondary structure comprises at least two α-helix structures.

16. 16. The recombinant single-chain meganuclease of item 14 wherein:

said stable secondary structure comprises from N-terminus to C-terminus a first loop, a first α-helix, a first turn, a second α-helix, and a second loop.

17. 17. The recombinant single-chain meganuclease of item 14 wherein:

said linker comprises 23-56 residues.

18. 18. A method for producing a genetically-modified eukaryotic cell including an exogenous sequence of interest inserted in a chromosome of said eukaryotic cell, comprising:

transfecting a eukaryotic cell with one or more nucleic acids including

1. (i) a first nucleic acid sequence encoding a meganuclease, and
2. (ii) a second nucleic acid sequence including said sequence of interest;

wherein said meganuclease produces a cleavage site in said chromosome and said sequence of interest is inserted into said chromosome at said cleavage site; and

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17.

19. 19. A method as in item 18 wherein:

said second nucleic acid further comprises sequences homologous to sequences flanking said cleavage site and said sequence of interest is inserted at said cleavage site by homologous recombination.

20. 20. A method as in item 18 wherein:

said second nucleic acid lacks substantial homology to said cleavage site and said sequence of interest is inserted into said chromosome by non-homologous end-joining.

21. 21. A method for producing a genetically-modified eukaryotic cell including an exogenous sequence of interest inserted in a chromosome of said eukaryotic cell, comprising:

introducing a meganuclease protein into a eukaryotic cell; and

transfecting said eukaryotic cell with a nucleic acid including said sequence of interest;

wherein said meganuclease produces a cleavage site in said chromosome and said sequence of interest is inserted into said chromosome at said cleavage site; and

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17.

22. 22. A method as in item 21 wherein:

said nucleic acid further comprises sequences homologous to sequences flanking said cleavage site and said sequence of interest is inserted at said cleavage site by homologous recombination.

23. 23. A method as in item 21 wherein:

said nucleic acid lacks substantial homology to said cleavage site and said sequence of interest is inserted into said chromosome by non-homologous end-joining.

24. 24. A method for producing a genetically-modified eukaryotic cell by disrupting a target sequence in a chromosome of said eukaryotic cell, comprising:

transfecting a eukaryotic cell with a nucleic acid encoding a meganuclease;

wherein said meganuclease produces a cleavage site in said chromosome and said target sequence is disrupted by non-homologous end-joining at said cleavage site; and

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17.

25. 25. A method of producing a genetically-modified organism comprising:

producing a genetically-modified eukaryotic cell according to the method of any one of items 18-24; and

growing said genetically-modified eukaryotic cell to produce said genetically-modified organism.

26. 26. A method as in item 25 wherein:

said eukaryotic cell is selected from the group consisting of a gamete, a zygote, a blastocyst cell, an embryonic stem cell, and a protoplast cell.

27. 27. A method for treating a disease by gene therapy in a eukaryote, comprising:

transfecting at least one cell of said eukaryote with one or more nucleic acids including

1. (i) a first nucleic acid sequence encoding a meganuclease, and
2. (ii) a second nucleic acid sequence including a sequence of interest;

wherein said meganuclease produces a cleavage site in said chromosome and said sequence of interest is inserted into said chromosome at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17; and

wherein insertion of said sequence of interest provides said gene therapy for said disease.

28. 28. A method as in item 27 wherein:

said second nucleic acid sequence further comprises sequences homologous to sequences flanking said cleavage site and said sequence of interest is inserted at said cleavage site by homologous recombination.

29. 29. A method as in item 27 wherein:

said second nucleic acid sequence lacks substantial homology to said cleavage site and said sequence of interest is inserted into said chromosome by non-homologous end-joining.

30. 30. A method for treating a disease by gene therapy in a eukaryote, comprising:

introducing a meganuclease protein into at least one cell of said eukaryote; and

transfecting said eukaryotic cell with a nucleic acid including a sequence of interest;

wherein said meganuclease produces a cleavage site in said chromosome and said sequence of interest is inserted into said chromosome at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17; and

wherein insertion of said sequence of interest provides said gene therapy for said disease.

31. 31. A method as in item 30 wherein:

said nucleic acid further comprises sequences homologous to sequences flanking said cleavage site and said sequence of interest is inserted at said cleavage site by homologous recombination.

32. 32. A method as in item 30 wherein:

said nucleic acid lacks substantial homology to said cleavage site and said sequence of interest is inserted into said chromosome by non-homologous end-joining.

33. 33. A method for treating a disease by gene therapy in a eukaryote by disrupting a target sequence in a chromosome of said eukaryotic cell, comprising:

transfecting at least one cell of said eukaryote with a nucleic acid encoding a meganuclease;

wherein said meganuclease produces a cleavage site in said chromosome and said target sequence is disrupted by non-homologous end-joining at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17; and

wherein disruption of said target sequence provides said gene therapy for said disease.

34. 34. A method for treating a viral pathogen infection in a eukaryotic host by disrupting a target sequence in a genome of said viral pathogen, comprising:

transfecting at least one infected cell of said eukaryotic host with a nucleic acid encoding a meganuclease;

wherein said meganuclease produces a cleavage site in said viral genome and said target sequence is disrupted by non-homologous end-joining at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17; and

wherein disruption of said target sequence provides treatment for said infection.

35. 35. A method for treating a viral pathogen infection in a eukaryotic host by disrupting a target sequence in a genome of said viral pathogen, comprising:

transfecting at least one infected cell of said eukaryotic host with a first nucleic acid encoding a meganuclease and a second nucleic acid;

wherein said meganuclease produces a cleavage site in said viral genome and said target sequence is disrupted by homologous recombination of said viral genome and said second nucleic acid at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17;

wherein said second nucleic acid comprises sequences homologous to sequences flanking said cleavage site; and

wherein disruption of said target sequence provides treatment for said infection.

36. 36. A method for treating a prokaryotic pathogen infection in a eukaryotic host by disrupting a target sequence in a genome of said prokaryotic pathogen, comprising:

transfecting at least cell of said prokaryotic pathogen infecting said eukaryotic host with a nucleic acid encoding a meganuclease;

wherein said meganuclease produces a cleavage site in said prokaryotic genome and said target sequence is disrupted by non-homologous end-joining at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17; and

wherein disruption of said target sequence provides treatment for said infection.

37. 37. A method for treating a prokaryotic pathogen infection in a eukaryotic host by disrupting a target sequence in a genome of said prokaryotic pathogen, comprising:

transfecting at least cell of said prokaryotic pathogen infecting said eukaryotic host with a first nucleic acid encoding a meganuclease and a second nucleic acid;

wherein said meganuclease produces a cleavage site in said prokaryotic genome and said target sequence is disrupted by homologous recombination of said prokaryotic genome and said second nucleic acid at said cleavage site;

wherein said meganuclease is a recombinant single-chain meganuclease of any one of items 1-17;

wherein said second nucleic acid comprises sequences homologous to sequences flanking said cleavage site; and

wherein disruption of said target sequence provides treatment for said infection.

SEQUENCE LISTING

[0150]

<110> PRECISION BIOSCIENCES, Inc.

<120> RATIONALLY-DESIGNED SINGLE-CHAIN MEGANUCLEASES WITH NON-PALINDROMIC RECOGNITION SEQUENCES

<130> 111-062T1

<150> EP 08 845 549.8
<151> 2008-10-31

<150> PCT/US200/082072
<151> 2008-10-31

<150> 61/001,247
<151> 2007-10-31

<160> 110

<170> PatentIn version 3.5

<210> 1
<211> 163
<212> PRT
<213> Chlamydomonas reinhardtii

<400> 1

```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln Ser
            20                  25                  30

Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln Lys
        35                  40                  45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
    50                  55                  60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
                85                  90                  95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
            100                 105                 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
        115                 120                 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
    130                 135                 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys

145                 150                 155                 160

Ser Ser Pro
```

<210> 2
<211> 170
<212> PRT
<213> Monomastix sp.

<400> 2
```
Met Thr Thr Lys Asn Thr Leu Gln Pro Thr Glu Ala Ala Tyr Ile Ala
1               5                   10                  15

Gly Phe Leu Asp Gly Asp Gly Ser Ile Tyr Ala Lys Leu Ile Pro Arg
            20                  25                  30

Pro Asp Tyr Lys Asp Ile Lys Tyr Gln Val Ser Leu Ala Ile Ser Phe
        35                  40                  45

Ile Gln Arg Lys Asp Lys Phe Pro Tyr Leu Gln Asp Ile Tyr Asp Gln
    50                  55                  60

Leu Gly Lys Arg Gly Asn Leu Arg Lys Asp Arg Gly Asp Gly Ile Ala
65                  70                  75                  80

Asp Tyr Thr Ile Ile Gly Ser Thr His Leu Ser Ile Ile Leu Pro Asp
                85                  90                  95

Leu Val Pro Tyr Leu Arg Ile Lys Lys Lys Gln Ala Asn Arg Ile Leu
            100                 105                 110

His Ile Ile Asn Leu Tyr Pro Gln Ala Gln Lys Asn Pro Ser Lys Phe
        115                 120                 125

Leu Asp Leu Val Lys Ile Val Asp Asp Val Gln Asn Leu Asn Lys Arg
    130                 135                 140

Ala Asp Glu Leu Lys Ser Thr Asn Tyr Asp Arg Leu Leu Glu Glu Phe
145                 150                 155                 160

Leu Lys Ala Gly Lys Ile Glu Ser Ser Pro
                165                 170
```

<210> 3
<211> 218
<212> PRT
<213> Chlamydomonas moewusii

<400> 3

```
Met Ser Asn Phe Ile Leu Lys Pro Gly Glu Lys Leu Pro Gln Asp Lys
1               5                   10                  15

Leu Glu Glu Leu Lys Lys Ile Asn Asp Ala Val Lys Lys Thr Lys Asn
            20              25                  30

Phe Ser Lys Tyr Leu Ile Asp Leu Arg Lys Leu Phe Gln Ile Asp Glu
        35              40                  45

Val Gln Val Thr Ser Glu Ser Lys Leu Phe Leu Ala Gly Phe Leu Glu
    50              55                  60

Gly Glu Ala Ser Leu Asn Ile Ser Thr Lys Lys Leu Ala Thr Ser Lys
65              70                  75                  80

Phe Gly Leu Val Val Asp Pro Glu Phe Asn Val Thr Gln His Val Asn
                85                  90                  95

Gly Val Lys Val Leu Tyr Leu Ala Leu Glu Val Phe Lys Thr Gly Arg
            100                 105                 110

Ile Arg His Lys Ser Gly Ser Asn Ala Thr Leu Val Leu Thr Ile Asp
            115                 120                 125

Asn Arg Gln Ser Leu Glu Glu Lys Val Ile Pro Phe Tyr Glu Gln Tyr
    130                 135                 140

Val Val Ala Phe Ser Ser Pro Glu Lys Val Lys Arg Val Ala Asn Phe
145                 150                 155                 160

Lys Ala Leu Leu Glu Leu Phe Asn Asn Asp Ala His Gln Asp Leu Glu
            165                 170                 175

Gln Leu Val Asn Lys Ile Leu Pro Ile Trp Asp Gln Met Arg Lys Gln
            180                 185                 190

Gln Gly Gln Ser Asn Glu Gly Phe Pro Asn Leu Glu Ala Ala Gln Asp
        195                 200                 205

Phe Ala Arg Asn Tyr Lys Lys Gly Ile Lys
    210                 215
```

<210> 4

<211> 194

<212> PRT

<213> Desulfurococcus mobilis

<400> 4

```
Met His Asn Asn Glu Asn Val Ser Gly Ile Ser Ala Tyr Leu Leu Gly
1               5                   10                  15

Leu Ile Ile Gly Asp Gly Gly Leu Tyr Lys Leu Lys Tyr Lys Gly Asn
```

```
                  20                   25                   30

Arg Ser Glu Tyr Arg Val Val Ile Thr Gln Lys Ser Glu Asn Leu Ile
        35                  40                  45

Lys Gln His Ile Ala Pro Leu Met Gln Phe Leu Ile Asp Glu Leu Asn
    50                  55                  60

Val Lys Ser Lys Ile Gln Ile Val Lys Gly Asp Thr Arg Tyr Glu Leu
65                  70                  75                  80

Arg Val Ser Ser Lys Lys Leu Tyr Tyr Tyr Phe Ala Asn Met Leu Glu
                85                  90                  95

Arg Ile Arg Leu Phe Asn Met Arg Glu Gln Ile Ala Phe Ile Lys Gly
            100                 105                 110

Leu Tyr Val Ala Glu Gly Asp Lys Thr Leu Lys Arg Leu Arg Ile Trp
        115                 120                 125

Asn Lys Asn Lys Ala Leu Leu Glu Ile Val Ser Arg Trp Leu Asn Asn
    130                 135                 140

Leu Gly Val Arg Asn Thr Ile His Leu Asp Asp His Arg His Gly Val
145                 150                 155                 160

Tyr Val Leu Asn Ile Ser Leu Arg Asp Arg Ile Lys Phe Val His Thr
                165                 170                 175

Ile Leu Ser Ser His Leu Asn Pro Leu Pro Pro Glu Arg Ala Gly Gly
                180                 185                 190

Tyr Thr
```

<210> 5

<211> 235

<212> PRT

<213> Saccharomyces cerevisiae

<400> 5

```
Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser
1               5                   10                  15

Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu
            20                  25                  30

Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg
        35                  40                  45
```

34

```
Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn
    50                  55                  60

Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu
65                  70                  75                  80

Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val
                85                  90                  95

Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu
            100                 105                 110

Ala Asn Leu Phe Ile Val Asn Asn Lys Lys Thr Ile Pro Asn Asn Leu
        115                 120                 125

Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp
    130                 135                 140

Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys Ser Ile
145                 150                 155                 160

Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Glu Tyr Leu Val
            165                 170                 175

Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys Ile Asn
            180                 185                 190

Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe
        195                 200                 205

Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr Lys Leu
    210                 215                 220

Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
225                 230                 235
```

<210> 6
<211> 254
<212> PRT
<213> Emericella nidulans

<400> 6
```
Met Ser Asp Leu Thr Tyr Ala Tyr Leu Val Gly Leu Phe Glu Gly Asp
1               5                   10                  15

Gly Tyr Phe Ser Ile Thr Lys Lys Gly Lys Tyr Leu Thr Tyr Glu Leu
            20                  25                  30

Gly Ile Glu Leu Ser Ile Lys Asp Val Gln Leu Ile Tyr Lys Ile Lys
        35                  40                  45
```

```
Lys Ile Leu Gly Ile Gly Ile Val Ser Phe Arg Lys Ile Asn Glu Ile
    50              55              60

Glu Met Val Ala Leu Arg Ile Arg Asp Lys Asn His Leu Lys Ser Phe
65              70              75              80

Ile Leu Pro Ile Phe Glu Lys Tyr Pro Met Phe Ser Asn Lys Gln Tyr
            85              90              95

Asp Tyr Leu Arg Phe Arg Asn Ala Leu Leu Ser Gly Ile Ile Ser Leu
        100             105             110

Glu Asp Leu Pro Asp Tyr Thr Arg Ser Asp Glu Pro Leu Asn Ser Ile
        115             120             125

Glu Ser Ile Ile Asn Thr Ser Tyr Phe Ser Ala Trp Leu Val Gly Phe
    130             135             140

Ile Glu Ala Glu Gly Cys Phe Ser Val Tyr Lys Leu Asn Lys Asp Asp
145             150             155             160

Asp Tyr Leu Ile Ala Ser Phe Asp Ile Ala Gln Arg Asp Gly Asp Ile
            165             170             175

Leu Ile Ser Ala Ile Arg Lys Tyr Leu Ser Phe Thr Thr Lys Val Tyr
            180             185             190

Leu Asp Lys Thr Asn Cys Ser Lys Leu Lys Val Thr Ser Val Arg Ser
        195             200             205

Val Glu Asn Ile Ile Lys Phe Leu Gln Asn Ala Pro Val Lys Leu Leu
    210             215             220

Gly Asn Lys Lys Leu Gln Tyr Leu Leu Trp Leu Lys Gln Leu Arg Lys
225             230             235             240

Ile Ser Arg Tyr Ser Glu Lys Ile Lys Ile Pro Ser Asn Tyr
            245             250
```

<210> 7
<211> 9
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 7
gaaactgtc 9

<210> 8
<211> 9
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 8
gacagtttc 9

<210> 9
<211> 9
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 9
caaaacgtc 9

<210> 10
<211> 9
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 10
gacgttttg 9

<210> 11
<211> 9
<212> DNA

36

<213> Monomastix sp.

<400> 11
cagaacgtc 9

<210> 12
<211> 9
<212> DNA
<213> Monomastix sp.

<400> 12
gacgttctg 9

<210> 13
<211> 9
<212> DNA
<213> Monomastix sp.

<400> 13
ggaactgtc 9

<210> 14
<211> 9
<212> DNA
<213> Monomastix sp.

<400> 14
gacagttcc 9

<210> 15
<211> 9
<212> DNA
<213> Chlamydomonas moewusii

<400> 15
ataacggtc 9

<210> 16
<211> 9
<212> DNA
<213> Chlamydomonas moewusii

<400> 16
gaccgttat 9

<210> 17
<211> 9
<212> DNA
<213> Chlamydomonas moewusii

<400> 17
ttcgctacc 9

<210> 18
<211> 9
<212> DNA
<213> Chlamydomonas moewusii

<400> 18
ggtagcgaa 9

<210> 19
<211> 5
<212> DNA
<213> Saccharomyces cerevisiae

<400> 19
taggg 5

<210> 20
<211> 5
<212> DNA
<213> Saccharomyces cerevisiae

<400> 20
cccta 5

<210> 21
<211> 9
<212> DNA
<213> Saccharomyces cerevisiae

<400> 21
taatgggac 9

<210> 22
<211> 9
<212> DNA
<213> Saccharomyces cerevisiae

<400> 22
gtcccatta 9

<210> 23
<211> 8
<212> DNA
<213> Desulfurococcus mobilis

<400> 23
gccggaac 8

<210> 24
<211> 8
<212> DNA
<213> Desulfurococcus mobilis

<400> 24
gttccggc 8

<210> 25
<211> 7
<212> DNA
<213> Desulfurococcus mobilis

<400> 25
aacggcc 7

<210> 26
<211> 7
<212> DNA
<213> Desulfurococcus mobilis

<400> 26
ggccgtt 7

<210> 27
<211> 8
<212> DNA
<213> Emericella nidulans

<400> 27
tttacaga 8

<210> 28
<211> 8
<212> DNA
<213> Emericella nidulans

<210> 28
tctgtaaa 8

<210> 29
<211> 9
<212> DNA
<213> Emericella nidulans

<400> 29
ctgaggagg 9

<210> 30
<211> 9
<212> DNA
<213> Emericella nidulans

<400> 30
cctcctcag

<210> 31
<211> 163
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 31
```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15

Val Asp Gly Asp Gly Ser Ile Lys Ala Gln Ile Lys Pro Glu Gln Asn
            20                  25                  30

Arg Lys Phe Lys His Arg Leu Glu Leu Thr Phe Gln Val Thr Gln Lys
            35                  40                  45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
    50                  55                  60

Gly Tyr Val Tyr Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
                85                  90                  95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
            100                 105                 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
        115                 120                 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
    130                 135                 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145                 150                 155                 160

Ser Ser Pro
```

<210> 32
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 32
aggcatctca ttagagatgc ct 22

<210> 33
<211> 22

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 33
aggcatctct aatgagatgc ct 22

<210> 34
<211> 22
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 34
gaaactgtct cacgacgttt tg 22

<210> 35
<211> 22
<212> DNA
<213> Chlamydomonas reinhardtii

<400> 35
caaaacgtcg tgagacagtt tc 22

<210> 36
<211> 163
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 36
```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln Cys
            20                  25                  30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln Lys
        35                  40                  45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
    50                  55                  60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
                85                  90                  95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
            100                 105                 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
        115                 120                 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
        130                 135                 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145                 150                 155                 160

Ser Ser Pro
```

<210> 37
<211> 163
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 37

```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser
            20                  25                  30

Tyr Lys Phe Lys His Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys
            35                  40                  45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
        50                  55                  60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
                85                  90                  95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
            100                 105                 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp

        115                 120                 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
        130                 135                 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145                 150                 155                 160

Ser Ser Pro
```

<210> 38
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 38
tgcggtgtc 9

<210> 39
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 39
gacaccgca 9

<210> 40
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 40
tgcggtgtcn nnngacaccg ca 22

41

<210> 41
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 41
tgcggtgtcn nnngacaccg ca 22

<210> 42
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 42
caggctgtc 9

<210> 43
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 43
gacagcctg 9

<210> 44
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 44
caggctgtcn nnngacagcc tg 22

<210> 45
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 45

caggctgtcn nnngacagcc tg 22

<210> 46
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 46
tgcggtgtca ttagacagcc tg 22

<210> 47
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 47
caggctgtct aatgacaccg ca 22

<210> 48
<211> 260
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 48

```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15


Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln Cys
            20                  25                  30


Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln Lys
        35                  40                  45


Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
    50                  55                  60


Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80


Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Met Leu Glu
                85                  90                  95


Arg Ile Arg Leu Phe Asn Met Arg Glu Phe Leu Leu Tyr Leu Ala Gly
            100                 105                 110
```

```
Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln
        115             120             125

Ser Tyr Lys Phe Lys His Arg Leu Arg Leu Glu Phe Gln Val Thr Gln
        130             135             140

Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
145             150             155             160

Val Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser
            165             170             175

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
            180             185             190

Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg
        195             200             205

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
        210             215             220

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
225             230             235             240

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
            245             250             255

Lys Ser Ser Pro
            260
```

<210> 49
<211> 260
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 49
```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5               10              15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser
            20              25              30

Tyr Lys Phe Lys His Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys
        35              40              45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
    50              55              60
```

```
Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65              70              75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Met Leu Glu
            85              90              95

Arg Ile Arg Leu Phe Asn Met Arg Glu Phe Leu Leu Tyr Leu Ala Gly
            100             105             110

Phe Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln
        115             120             125

Cys Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln
    130             135             140

Lys Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly
145             150             155             160

Val Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser
            165             170             175

Glu Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu
            180             185             190

Lys Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg
        195             200             205

Leu Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr
    210             215             220

Trp Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr
225             230             235             240

Thr Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys
            245             250             255

Lys Ser Ser Pro
            260
```

<210> 50
<211> 334
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 50
```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5               10              15
```

45

```
Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln Cys
              20              25              30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln Lys
          35              40              45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
      50              55              60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65              70              75              80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
              85              90              95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
          100             105             110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
          115             120             125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
          130             135             140

Ser Glu Thr Val Arg Ala Val Gly Ser Ser Gly Ser Ser Gly Ser Ser
145             150             155             160

Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
              165             170             175

Ser Ser Gly Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly
              180             185             190

Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser Tyr Lys Phe Lys His
          195             200             205

Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp
      210             215             220

Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp
225             230             235             240

Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His
              245             250             255

Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln
          260             265             270

Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu Pro Ser Ala Lys Glu
          275             280             285

Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala
      290             295             300

Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg
305             310             315             320

Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
              325             330
```

<210> 51

<211> 350

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 51

```
Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1               5                   10                  15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Asp Pro Arg Gln Cys
            20                  25                  30

Arg Lys Phe Lys His Glu Leu Arg Leu Arg Phe Gln Val Thr Gln Lys
            35                  40                  45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
        50                  55                  60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65                  70                  75                  80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
                85                  90                  95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu
            100                 105                 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
            115                 120                 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
    130                 135                 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Pro Gly Ser Val Gly
145                 150                 155                 160

Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser
            165                 170                 175

Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala Gly Ala Thr Lys
            180                 185                 190

Ser Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly
            195                 200                 205

Ser Ile Ile Ala Gln Ile Lys Pro Glu Gln Ser Tyr Lys Phe Lys His
    210                 215                 220

Arg Leu Arg Leu Glu Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp
225                 230                 235                 240

Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp
            245                 250                 255

Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His
            260                 265                 270

Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln
            275                 280                 285

Ala Asn Leu Val Leu Lys Ile Ile Trp Arg Leu Pro Ser Ala Lys Glu
    290                 295                 300

Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala
305                 310                 315                 320

Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg
            325                 330                 335

Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
            340                 345                 350
```

<210> 52
<211> 373
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 52

```
Met Thr Thr Lys Asn Thr Leu Gln Pro Thr Glu Ala Ala Tyr Ile Ala
1               5                   10                  15

Gly Phe Leu Asp Gly Asp Gly Ser Ile Tyr Ala Lys Leu Ile Pro Arg
            20                  25                  30

Pro Asp Tyr Lys Asp Ile Lys Tyr Gln Val Ser Leu Ala Ile Ser Phe
            35                  40                  45

Ile Gln Arg Lys Asp Lys Phe Pro Tyr Leu Gln Asp Ile Tyr Asp Gln
        50                  55                  60

Leu Gly Lys Arg Gly Asn Leu Arg Lys Asp Arg Gly Asp Gly Ile Ala
65                  70                  75                  80

Asp Tyr Thr Ile Ile Gly Ser Thr His Leu Ser Ile Ile Leu Pro Asp
                85                  90                  95

Leu Val Pro Tyr Leu Arg Ile Lys Lys Lys Gln Ala Asn Arg Ile Leu
            100                 105                 110

His Ile Ile Asn Leu Tyr Pro Gln Ala Gln Lys Asn Pro Ser Lys Phe
            115                 120                 125

Leu Asp Leu Val Lys Ile Val Asp Asp Val Gln Asn Leu Asn Lys Arg
    130                 135                 140

Ala Asp Glu Leu Lys Ser Thr Asn Tyr Asp Arg Leu Leu Glu Glu Phe
145                 150                 155                 160

Leu Lys Ala Gly Lys Ile Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser
                165                 170                 175

Ala Ala Ser Ser Ala Ser Ser Ala Ala Ser Ser Pro Gly Ser Gly Pro
            180                 185                 190

Ser Glu Ala Leu Arg Ala Ala Ser Ser Phe Ala Ser Lys Pro Gly Ser
            195                 200                 205

Thr Leu Gln Pro Thr Glu Ala Ala Tyr Ile Ala Gly Phe Leu Asp Gly
    210                 215                 220

Asp Gly Ser Ile Tyr Ala Lys Leu Ile Pro Arg Pro Asp Tyr Lys Asp
225                 230                 235                 240

Ile Lys Tyr Gln Val Ser Leu Ala Ile Ser Phe Ile Gln Arg Lys Asp
            245                 250                 255

Lys Phe Pro Tyr Leu Gln Asp Ile Tyr Asp Gln Leu Gly Lys Arg Gly
        260                 265                 270

Asn Leu Arg Lys Asp Arg Gly Asp Gly Ile Ala Asp Tyr Thr Ile Ile
        275                 280                 285

Gly Ser Thr His Leu Ser Ile Ile Leu Pro Asp Leu Val Pro Tyr Leu

    290                 295                 300

Arg Ile Lys Lys Lys Gln Ala Asn Arg Ile Leu His Ile Ile Asn Leu
305                 310                 315                 320

Tyr Pro Gln Ala Gln Lys Asn Pro Ser Lys Phe Leu Asp Leu Val Lys
                325                 330                 335

Ile Val Asp Asp Val Gln Asn Leu Asn Lys Arg Ala Asp Glu Leu Lys
            340                 345                 350

Ser Thr Asn Tyr Asp Arg Leu Leu Glu Glu Phe Leu Lys Ala Gly Lys
            355                 360                 365

Ile Glu Ser Ser Pro
    370
```

<210> 53

<211> 22

<212> DNA

<213> Monomastix sp.

<400> 53

ggaactgtct cacgacgttc tg 22

<210> 54

48

<211> 22
<212> DNA
<213> Monomastix sp.

<400> 54
cagaacgtcg tgagacagtt cc 22

<210> 55
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 55
Leu Ala Gly Leu Ile Asp Ala Asp Gly
1   5

<210> 56
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 56
tgcggtgtcn nnngacagcc tg 22

<210> 57
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<220>
<221> modified_base
<222> (10)..(13)
<223> a, c, t, g, u, unknown or other

<400> 57
caggctgtcn nnngacaccg ca 22

<210> 58
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 58
Ser Gly Gly Ser
1

<210> 59
<211> 22
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

49

<400> 59
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15

Ser Ser Gly Ser Ser Gly
            20
```

<210> 60
<211> 25
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 60
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly

1               5                   10                  15

Ser Ser Gly Ser Ser Gly Ser Ser Gly
            20                  25
```

<210> 61
<211> 28
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 61
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15

Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
            20                  25
```

<210> 62
<211> 31
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 62
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15

Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
            20                  25                  30
```

<210> 63
<211> 34
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 63
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15

Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser
            20                  25                  30

Ser Gly
```

<210> 64
<211> 29
<212> PRT
<213> Artificial Sequence

50

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 64
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15


Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Gly
            20                  25
```

<210> 65
<211> 30
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 65
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15


Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Gly
            20                  25                  30
```

<210> 66
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 66
```
Val Leu Asp Ser Pro Gly Ser Val
1               5
```

<210> 67
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 67
```
Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser
1               5                   10
```

<210> 68
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 68
```
Leu Ser Pro Ser Gln Ala
1               5
```

<210> 69
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 69
```
Ala Ser Ser Ser Pro Gly Ser Gly Ile
1               5
```

<210> 70

<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 70
Ala Ser Ser Ser
1

<210> 71
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 71
Pro Gly Ser Gly Ile
1   5

<210> 72
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 72
Ile Ser Glu Ala Leu Arg
1   5

<210> 73
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 73
Ser Glu Ala Leu Arg Ala
1   5

<210> 74
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 74
Ala Ser Ser Ala
1

<210> 75
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 75
Ala Leu Arg Ala Gly Ala
1   5

<210> 76
<211> 6

52

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 76
```
Thr Lys Ser Lys Glu Phe
1               5
```

<210> 77
<211> 40
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 77
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20                  25                  30

Leu Arg Ala Gly Ala Thr Lys Ser
        35                  40
```

<210> 78
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 78
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20                  25                  30

Leu Arg Ala Gly Gly Ala Thr Lys Ser
        35                  40
```

<210> 79
<211> 42
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 79
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20                  25                  30

Leu Arg Ala Ala Gly Gly Ala Thr Lys Ser
        35                  40
```

<210> 80
<211> 43
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 80

```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20              25                  30

Leu Arg Ala Ala Ser Gly Gly Ala Thr Lys Ser
        35                  40
```

<210> 81
<211> 44
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 81
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20              25                  30

Leu Arg Ala Ala Ser Ser Gly Gly Ala Thr Lys Ser
        35                  40
```

<210> 82
<211> 45
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 82
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20              25                  30

Leu Arg Ala Ala Ser Ser Ala Gly Gly Ala Thr Lys Ser
        35                  40                  45
```

<210> 83
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 83
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala
            20              25                  30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 84
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 84

```
Ser Leu Pro Gly Ser Val Gly Gly Ile Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala
            20                  25                  30

Pro Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 85
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 85
```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala
            20                  25                  30

Thr Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 86
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 86
```
Ser Leu Pro Gly Ser Leu Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Pro Ser Glu Ala
            20                  25                  30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 87
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 87
```
Ser Leu Pro Gly Ser Val Gly Gly Val Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Val Ser Glu Ala
            20                  25                  30

Ser Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 88
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 88

```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Leu Ser Glu Ala
            20                  25                  30

Leu Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 89
<211> 41
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 89
```
Ser Leu Pro Gly Ser Leu Gly Gly Ile Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ser Ser Glu Ala
            20                  25                  30

Ser Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 90
<211> 40
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 90
```
Ser Pro Gly Ser Val Gly Gly Val Ser Pro Ser Gln Ala Ser Ser Ala
1               5                   10                  15

Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Thr
            20                  25                  30

Arg Ala Gly Ala Thr Lys Glu Phe
        35                  40
```

<210> 91
<211> 37
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 91
```
Ser Leu Pro Gly Ser Leu Gly Gly Val Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Pro Gly Ser Gly Thr Ser Glu Ala Pro Arg Ala Gly
            20                  25                  30

Ala Thr Lys Glu Phe
        35
```

<210> 92
<211> 37
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 92

56

```
Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser
1               5                   10                  15

Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Ile Arg Ala Gly
            20              25                  30

Ala Thr Lys Glu Phe
            35
```

<210> 93
<211> 44
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 93
```
Ser Leu Pro Gly Ser Leu Gly Gly Val Ser Pro Ser Gln Ala Ser Ser

1               5                   10                  15

Ala Ala Ser Ser Ala Ser Ser Ala Ala Ser Ser Pro Gly Ser Gly Ala
            20                  25                  30

Ser Glu Ala Ser Arg Ala Gly Ala Thr Lys Glu Phe
            35                  40
```

<210> 94
<211> 6
<212> PRT
<213> Monomastix sp.

<400> 94
```
Leu Gln Pro Thr Glu Ala
1               5
```

<210> 95
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 95
```
Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala
1               5                   10                  15

Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg
            20                  25                  30

Ala Gly Ala Thr Lys Ser Ala
            35
```

<210> 96
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 96
```
Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser Ser Ala Ala
1               5                   10                  15

Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg
            20                  25                  30

Ala Gly Ala Thr Lys Ser Gly
            35
```

<210> 97
<211> 43

57

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 97
```
Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
1               5                   10                  15

Ser Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala
            20                  25                  30

Ala Ser Ser Leu Ala Ser Lys Pro Gly Ser Thr
        35                  40
```

<210> 98
<211> 39
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 98
```
Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
1               5                   10                  15

Ser Ala Ala Ser Ser Pro Gly Ser Gly Ile Ser Glu Ala Leu Arg Ala
            20                  25                  30

Ala Ser Ser Pro Gly Ser Thr
        35
```

<210> 99
<211> 43
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 99
```
Gly Gly Ala Ser Pro Ser Gln Ala Ser Ser Ala Ala Ser Ser Ala Ser
1               5                   10                  15

Ser Ala Ala Ser Ser Pro Gly Ser Gly Pro Ser Glu Ala Leu Arg Ala
            20                  25                  30

Ala Ser Ser Phe Ala Ser Lys Pro Gly Ser Thr
        35                  40
```

<210> 100
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 100
```
Leu Pro Gly Glu
1
```

<210> 101
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 101
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10
```

<210> 102
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 102
```
Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser Gly
1               5                   10                  15
```

<210> 103
<211> 29
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<400> 103
```
Asn Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln Ala Ser
1               5                   10                  15

Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly
                20                  25
```

<210> 104
<211> 11
<212> PRT
<213> Desulfurococcus mobilis

<400> 104
```
Met Leu Glu Arg Ile Arg Leu Phe Asn Met Arg
1               5                   10
```

<210> 105
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 105
tgcggtgtca ttagacaccg ca 22

<210> 106
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 106
tgcggtgtct aatgacaccg ca 22

<210> 107
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 107
caggctgtca ttagacagcc tg 22

<210> 108
<211> 22

59

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 108
caggctgtca ttagacagcc tg 22

<210> 109
<211> 46
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 109
```
Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
1               5                   10                  15


Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
            20                  25                  30


Ser Glu Ala Leu Arg Ala Gly Ala Thr Lys Ser Lys Glu Phe
        35                  40                  45
```

<210> 110
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic 6xHis tag

<400> 110
```
His His His His His His
1               5
```

REFERENCES CITED IN THE DESCRIPTION

Patent documents cited in the description

- WO2008010009A [0007]
- WO2007093918A [0007]
- WO2007093836A [0007]
- WO2006097764A [0007]
- WO2006059317A [0007]
- WO2008059382A [0007]
- WO2008102198A [0007]
- WO2007060495A [0007]
- WO2007049156A [0007]
- WO2006097853A [0007]
- WO2004067736A [0007]
- WO2007047859A [0007] [0008] [0008] [0021] [0071] [0077] [0093] [0096] [0102] [0105] [0105] [0105] [0145] [0146] [0146] [0147]
- WO2006097654A [0008]
- WO2007057781A [0008]
- WO2007049095A [0008]
- WO2007034262A [0008]

- WO2006093249A [0008]
- WO2006093152A [0008]
- WO2003078619A [0010] [0011] [0042] [0070] [0072] [0131] [0132] [0134]
- US7037492B [0111]
- US5384253A [0112]
- US4873316A [0116]
- EP0264166A [0116]
- US6316024B [0124]
- US6379699B [0124]
- US6387397B [0124]
- US6511676B [0124]
- US6593308B [0124]
- EP08845549A [0150]
- US200082072W [0150]
- WO61001247A [0150]

Non-patent literature cited in the description

- **PORTEUS et al.**Nat. Biotechnol., 2005, vol. 23, 967-73 [0002] [0003] [0003] [0006]
- **TZFIRA et al.**Trends Biotechnol, 2005, vol. 23, 567-9 [0002]
- **MCDANIEL et al.**Curr. Opin. Biotechnol., 2005, vol. 16, 476-83 [0002]
- **PORTEUS**Mol. Ther, 2006, vol. 13, 438-46 [0003]
- **WRIGHT et al.**Plant J., 2005, vol. 44, 693-705 [0003]
- **URNOV et al.**Nature, 2005, vol. 435, 646-51 [0003]
- **SMITH et al.**Nucleic Acids Res., 2000, vol. 28, 3361-9 [0003]
- **STODDARD**Q. Rev. Biophys., 2006, vol. 38, 49-95 [0004] [0004]
- **CHEVALIER et al.**Nucleic Acids Res., 2001, vol. 29, 183757-3774 [0004]
- **MONNAT et al.**Biochem. Biophys. Res. Common., 1999, vol. 255, 88-93 [0005]
- **ROUET et al.**Mol. Cell. Biol., 1994, vol. 14, 8096-106 [0005]
- **CHILTON et al.**Plant Physiol., 2003, vol. 133, 956-65 [0005]
- **PUCHTA et al.**Proc. Natl. Acad. Sci. USA, 1996, vol. 93, 5055-60 [0005]
- **RONG et al.**Genes Dev., 2002, vol. 16, 1568-81 [0005]
- **GOUBLE et al.**J. Gene Med, 2006, vol. 8, 5616-622 [0005]
- **SUSSMAN et al.**J. Mol. Biol., 2004, vol. 342, 31-41 [0006] [0007] [0104] [0105]
- **EPINAT et al.**Nucleic Acids Res., 2003, vol. 31, 2952-62 [0006] [0010] [0011] [0042] [0070] [0131] [0132] [0134]
- **CHAMES et al.**Nucleic Acids Res., 2005, vol. 33, e178- [0007]
- **SELIGMAN et al.**Nucleic Acids Res., 2002, vol. 30, 3870-9 [0007]
- **ARNOULD et al.**J. Mol. Biol., 2006, vol. 355, 443-58 [0007] [0104] [0105]
- **ROSEN et al.**Nucleic Acids Res., 2006, vol. 34, 4791-4800 [0007]
- **ARNOULD et al.**J. Mol. Biol., 2007, vol. 371, 49-65 [0007]
- **SMITH et al.**Nuc. Acids Res., 2006, vol. 34, 149-157 [0008]
- **CHAMES et al.**Nucleic Acids Res., 2005, vol. 33, 178-186 [0008]
- **FAJARDO-SANCHEZ et al.**Nucleic Acids Res., 2008, vol. 36, 2163-2173 [0008] [0011] [0070] [0133]
- **CHEVALIER et al.**Mol. Cell, 2002, vol. 10, 895-905 [0010] [0042] [0103]
- **CAHILL et al.**Front. Biosci, 2006, vol. 11, 1958-1976 [0063] [0064]
- **SALOMON et al.**EMBD J., 1998, vol. 17, 6086-6095 [0064]
- **AUSUBEL et al.**Current Protocols in Molecular BiologyWiley19990000 [0065] [0128]
- **ALTSCHUL et al.**J. Mol. Biol., 1990, vol. 215, 403-410 [0066]
- **GISHSTATES**Nature Genet, 1993, vol. 3, 266-272 [0066]
- **MADDEN et al.**Meth. Enzymol, 1996, vol. 266, 131-141 [0066]
- **ALTSCHUL et al.**Nucleic Acids Res., 1997, vol. 25, 33 893402- [0066]
- **ZHANG et al.**J. Comput. Biol., 2000, vol. 7, 1-2203-14 [0066]
- **PRIETO et al.**Nucl. Acids Res., 2007, vol. 35, 3262-3271 [0072]
- **EPINAT et al.**Nucl. Acids Res, 2003, vol. 31, 2952-62 [0072]
- **JURICA et al.**Mol. Cell, 1998, vol. 2, 469-476 [0073] [0133]
- **AURORAROSE**Protein Sci., 1998, vol. 7, 21-38 [0076] [0090] [0090] [0090]
- **FERSHT**Structure and Mechanism in Protein ScienceW.H. Freeman19980000 [0076] [0063] [0090]
- **MACK et al.**Proc. Nat. Acad. Sci., 1995, vol. 92, 7021-7025 [0083]
- **UEDA et al.**J. Immunol. Methods, vol. 241, 159-170 [0083]

- **KIM et al.**Proc. Nat. Acad. Sci., 1996, vol. 93, 1156-1160 [0083]
- **SPIEGEL et al.**Structure, 2006, vol. 14, 869-880 [0098]
- **CHAMES et al.**Nucl. Acids Res., 2005, vol. 33, e178- [0104] [0105]
- **SELIGMAN et al.**Nucl. Acids Res., 2002, vol. 30, 3870-9 [0104] [0105]
- **ASHWORTH et al.**Nature, 2006, vol. 441, 656-659 [0105]
- **ARNOULD et al.**J. Mol. Biol, 2007, vol. 371, 149-65 [0105]
- **HUDECZ et al.**Med. Res. Rev., 2005, vol. 25, 679-736 [0109]
- **AUSUBEL**Current Protocols in Molecular BiologyWiley19990000 [0110]
- **LASIC et al.**Science, 1995, vol. 267, 1275-76 [0111] [0124] [0124]
- **RUI et al.**Life Sci., 2002, vol. 71, 151771-8 [0111]
- **GRAHAM et al.**Virology, 1973, vol. 54, 2536-539 [0112]
- **ZATLOUKAL et al.**Ann. N.Y. Acad. Sci., 1992, vol. 660, 136-153 [0112]
- **CAPECCHI**Cell, 1980, vol. 22, 2479-488 [0112]
- **WONG et al.**Biochim. Biophys. Res. Commun., 1982, vol. 107, 2584-587 [0112]
- **FROMM et al.**Proc. Nat'l Acad. Sci. USA, 1985, vol. 82, 175824-5828 [0112]
- **JOHNSTON et al.**Methods Cell. Biol., 1994, vol. 43, A353-365 [0112]
- **FYNAN et al.**Proc. Nat'l Acad. Sci. USA, 1993, vol. 90, 2411478-11482 [0112]
- **CLAPP**Clin. Perinatol., 1993, vol. 20, 1155-168 [0112]
- **LU et al.**J. Exp. Med., 1993, vol. 178, 62089-2096 [0112]
- **EGLITIS et al.**Avd. Exp. Med. Biol., 1988, vol. 241, 19-27 [0112]
- **EGLITIS et al.**Biotechniques, 1988, vol. 6, 7608-614 [0112]
- **CURIEL et al.**Proc. Nat'l Acad. Sci. USA, 1991, vol. 88, 198850-8854 [0112]
- **CURIEL et al.**Hum. Gen. Ther., 1992, vol. 3, 2147-154 [0112]
- **WAGNER et al.**Proc. Nat'l Acad. Sci. USA, 1992, vol. 89, 136099-6103 [0112]
- **GREMILLON et al.**Plant J., 2004, vol. 37, 218-228 [0113]
- **SINGH et al.**FEBS Lett, 2003, vol. 542, 47-52 [0113]
- **OMIRULLEH et al.**Plant Molecular Biology, 1993, vol. 21, 415-428 [0114]
- **GOSSLER et al.**Pro. Natl. Acad. Sci. USA, 1986, vol. 83, 9065-9069 [0115]
- **PINKERT et al.**Genes Dev., 1987, vol. 1, 268-277 [0116]
- **CALAMEEATON**Adv. Immunol., 1988, vol. 43, 235-275 [0116]
- **WINOTOBALTIMORE**EMBO J., 1989, vol. 8, 729-733 [0116]
- **BANERJI et al.**Cell, 1983, vol. 33, 729-740 [0116]
- **QUEENBALTIMORE**Cell, 1983, vol. 33, 741-748 [0116]
- **BYRNERUDDLE**Proc. Natl. Acad. Sci. USA, 1989, vol. 86, 5473-5477 [0116]
- **EDLUND et al.**Science, 1985, vol. 230, 912-916 [0116]
- **KESSELGRUSS**Science, 1990, vol. 249, 374-379 [0116]
- **CAMPESTILGHMAN**Genes Dev., 1989, vol. 3, 537-546 [0116]
- **KORMAN et al.**Proc. Nat'l Acad. Sci., 1987, vol. 84, 2150-2154 [0123]
- **YOUNG et al.**J. Calif. Dent. Assoc., 2005, vol. 33, 12967-71 [0124]
- **PFEIFFER et al.**J. Valc. Surg., 2006, vol. 43, 51021-7 [0124]
- **VAN DER GIESSEN et al.**Microbiology, 1994, vol. 140, 1103-1108 [0129]
- **CHEVALIER et al.**Biochemistry, 2001, vol. 43, 14015-14026 [0130]
- **PAPWORTH et al.**Gene, 2006, vol. 366, 27-38 [0130]
- **SANGER et al.**Proc. Natl. Acad. Sci. USA., 1977, vol. 74, 125463-7 [0135]

1

Patentkrav

1.   Rekombinant enkeltkædet meganuklease, som omfatter:

en første LAGLIDADG-subunit, der omfatter en polypeptidsekvens, som udviser

5   mindst 85% sekvensidentitet med resterne 9-151 i en vildtype-I-CreI-meganukle-

ase ifølge SEQ ID NO: 1 og har et første genkendelseshalvsted;

en anden LAGLIDADG-subunit, der omfatter en polypeptidsekvens, som udviser

mindst 85% sekvensidentitet med resterne 9-151 i en vildtype-I-CreI-meganukle-

ase ifølge SEQ ID NO: 1 og har et andet genkendelseshalvsted;

10   hvor den første og den anden LAGLIDADG-subunit er kovalent forbundet med

en polypeptidlinker, hvor linkeren er en fleksibel linker og omfatter mere end 25

og mindre end 31 rester;

idet den første subunit er kovalent bundet til polypeptidlinkeren ved en rest sva-

rende til en position valgt fra gruppen bestående af positionerne 151-153 ifølge

15   SEQ ID NO: 1;

idet den anden subunit er kovalent bundet til polypeptidlinkeren ved en rest sva-

rende til en position valgt fra gruppen bestående af positionerne 7-9 ifølge SEQ

ID NO: 1; og

hvor den første og den anden LAGLIDADG-subunit kan fungere sammen til gen-

20   kendelse og spaltning af en ikke-palindromisk DNA-sekvens, som er en hybrid af

det første genkendelseshalvsted og det andet genkendelseshalvsted.

2.   Rekombinant enkeltkædet meganuklease, som omfatter:

en første LAGLIDADG-subunit, der omfatter en polypeptidsekvens, som udviser

25   mindst 85% sekvensidentitet med resterne 9-151 i en vildtype-I-CreI-meganukle-

ase ifølge SEQ ID NO: 1 og har et første genkendelseshalvsted;

en anden LAGLIDADG-subunit, der omfatter en polypeptidsekvens, som udviser

mindst 85% sekvensidentitet med resterne 9-151 i en vildtype-I-CreI-meganukle-

ase ifølge SEQ ID NO: 1 og har et andet genkendelseshalvsted;

30   hvor den første og den anden LAGLIDADG-subunit er kovalent forbundet med

en polypeptidlinker,

2

hvor linkeren fra den N-terminale ende til den C-terminale ende omfatter en første sløjfe, en første α-helix, en første drejning, en anden α-helix og en anden sløjfe, eller

hvor linkeren består af en hvilken som helst af SEQ ID NO: 77-93,

5   idet den første subunit er kovalent bundet til polypeptidlinkeren ved en rest svarende til en position valgt fra gruppen bestående af positionerne 152-163 ifølge SEQ ID NO: 1;

idet den anden subunit er kovalent bundet til polypeptidlinkeren ved en rest svarende til en position valgt fra gruppen bestående af positionerne 1-9 ifølge SEQ

10   ID NO: 1; og

hvor den første og den anden LAGLIDADG-subunit kan fungere sammen til genkendelse og spaltning af en ikke-palindromisk DNA-sekvens, som er en hybrid af det første genkendelseshalvsted og det andet genkendelseshalvsted.

15   3.   Rekombinant enkeltkædet meganuklease ifølge krav 1-2, hvor mindst ét af LAGLIDADG-domænerne omfatter mindst én aminosyremodifikation valgt fra gruppen bestående af: Y75, L75, C75, Y139, C46, A46, H75, R75, H46, K46, R46, K70, E70, E75, E46, D46, Q70, C70, L70, Q75, H139, Q46, G70, A70, S70, G46, T44, A44, V44, I44, L44, N44, D70, K44, R44, H70, D44, E44, C44, Q68,

20   C24, E68, F68, K24, R24, M68, C68, L68, H68, Y68, K68, A26, Q77, E77, K26, R77, E26, S77, Q26, S26, E42, R42, K28, C28, Q42, M66, K66, Q40, E40, R28, R40, C40, I40, V40, C79, I79, V79, Q28, A40, A79, A28, H28, S40, S28, E38, K30, R30, K38, R38, E30, I38, L38, C38, H38, N38, Q30, F33, E33, D33, H33, L33, V33, I33, C33, R32, R33, E32, K32, L32, V32, A32, C32, D32, I32, N32,

25   H32, Q32 og T32.

4.   Rekombinant enkeltkædet meganuklease ifølge krav 1-3, hvor hver af LAG-LIDADG-subunitterne har et genkendelseshalvsted valgt fra gruppen bestående af SEQ ID NO: 7-30.

30

5.   Rekombinant enkeltkædet meganuklease ifølge krav 4, hvor mindst én af LAGLIDADG-subunitterne har et genkendelseshalvsted valgt fra gruppen bestå-

ende af SEQ ID NO: 7-30; og den anden LAGLIDADG-subunit har et genkendel-seshalvsted, som afviger med en modifikation af mindst ét basepar fra et gen-kendelseshalvsted valgt fra gruppen bestående af SEQ ID NO: 7-30.

5   6.   Rekombinant enkeltkædet meganuklease ifølge krav 2-5, hvor linkeren har en stabil sekundær struktur og omfatter 23-56 rester.

7.   Rekombinant enkeltkædet meganuklease ifølge krav 1, hvor mindst 50% af de rester, der udgør linkeren, er polære uladede rester.

10

8.   *Ex vivo*-fremgangsmåde til frembringelse af en genmodificeret eukaryot celle, som indbefatter en eksogen sekvens af interesse indsat i et kromosom i den eukaryote celle, hvilken fremgangsmåde omfatter:

(a) transfektion af en eukaryot celle med en eller flere nukleinsyrer, som indbe-
15   fatter
(i) en første nukleinsyresekvens, der koder for en meganuklease, og
(ii) en anden nukleinsyresekvens, som indbefatter sekvensen af interesse;
hvor meganukleasen frembringer et spaltningssted i kromosomet, og sekvensen af interesse indsættes i kromosomet på spaltningsstedet; og hvor meganuklea-
20   sen er en rekombinant enkeltkædet meganuklease ifølge et hvilket som helst af kravene 1-7, eller
(b) indføring af et meganukleaseprotein i en eukaryot celle og transfektion af den eukaryote celle med en nukleinsyre, som indbefatter sekvensen af interesse;
hvor meganukleasen frembringer et spaltningssted i kromosomet, og sekvensen
25   af interesse indsættes i kromosomet på spaltningsstedet; og hvor meganuklea-sen er en rekombinant enkeltkædet meganuklease ifølge et hvilket som helst af kravene 1-7.

9.   *Ex vivo*-fremgangsmåde til frembringelse af en genmodificeret eukaryot
30   celle ved at introducere forstyrrelse i en målsekvens i et kromosom i den euka-ryote celle, hvilken fremgangsmåde omfatter: transfektion af en eukaryot celle med en nukleinsyre, som koder for en meganuklease, hvor meganukleasen frem-

4

bringer et spaltningssted i kromosomet, og der introduceres forstyrrelse i målse-kvensen ved hjælp af ikke-homolog endeforbindelse på spaltningsstedet, og hvor meganukleasen er en rekombinant enkeltkædet meganuklease ifølge et hvilket som helst af kravene 1-7.

5

10. Fremgangsmåde til frembringelse af en ikke-human genmodificeret orga-nisme, hvilken fremgangsmåde omfatter: frembringelse af en genmodificeret eu-karyot celle ved fremgangsmåden ifølge et hvilket som helst af kravene 8-9, hvor den eukaryote celle er en protoplast, en ikke-human gamet, en ikke-human

10 zygote, en ikke-human blastocyst eller en ikke-human embryonal stamcelle, og dyrkning af den genmodificerede eukaryote celle for at frembringe den genmodi-ficerede organisme.

11. Fremgangsmåde ifølge krav 10, hvor den ikke-humane modificerede orga-

15 nisme er en plante, og den genmodificerede eukaryote celle er en protoplast.

12. Fremgangsmåde ifølge krav 10, hvor den ikke-humane genmodificerede or-ganisme er et dyr, og den genmodificerede eukaryote celle er en ikke-human zygote eller en ikke-human embryonal stamcelle.

20

13. Rekombinant enkeltkædet meganuklease ifølge et hvilket som helst af kra-vene 1-7 til anvendelse til genterapi.

14. Rekombinant enkeltkædet meganuklease ifølge et hvilket som helst af kra-

25 vene 1-7 til anvendelse til behandling af en viral patogeninfektion hos en eukaryot vært eller en prokaryot patogeninfektion hos en eukaryot vært.

DRAWINGS

Figure 1

Structural Components of Linker 9