

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 May 2009 (07.05.2009)

PCT

(10) International Publication Number  
**WO 2009/059195 A2**

(51) International Patent Classification:

C12N 9/22 (2006.01)

(21) International Application Number:

PCT/US2008/082072

(22) International Filing Date: 31 October 2008 (31.10.2008)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/001,247

31 October 2007 (31.10.2007) US

(71) Applicant (for all designated States except US): **PRECISION BIOSCIENCES** [—/US]; 104 T.w. Alexander Drive, Bldg. 7, Research Triangle Park, NC 27709 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SMITH, James, Jefferson** [US/US]; 1808 Carnation Drive, Durham, NC 27703 (US). **JANTZ, Derek** [US/US]; 215 Hillview Drive, Durham, NC 27703 (US).

(74) Agents: **TWOMEY, Michael, J.** et al.; Wilmer Cutler Pickering Hale And Dorr LLP, 60 State Street, Boston, MA 02109 (US).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every

kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

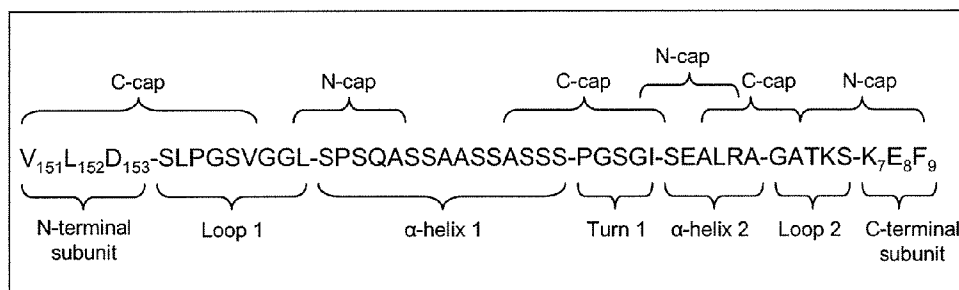
Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: RATIONALLY-DESIGNED SINGLE-CHAIN MEGANUCLEASES WITH NON-PALINDROMIC RECOGNITION SEQUENCES

Figure 1

### Structural Components of Linker 9



(57) Abstract: Disclosed are 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.

## **RATIONALLY-DESIGNED SINGLE-CHAIN MEGANUCLEASES WITH NON-PALINDROMIC RECOGNITION SEQUENCES**

### **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. Commun.* 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 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, *inter alia*, genetic engineering, gene therapy, treatment of pathogenic infections, and *in vitro* applications in diagnostics and research.

[0014] Thus, in some embodiments, the invention provides 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. In some embodiments, the mono-LAGLIDADG subunit is derived from a wild-type meganuclease selected from I-CreI, I-MsoI and I-CeuI.

[0015] In other embodiments, the invention provides 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-CreI, I-MsoI and I-CeuI, but the N-terminal subunit is derived from a wild-type meganuclease of a different species than the C-terminal subunit.

**[0016]** In some embodiments, the invention provides 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-CreI, I-MsoI and I-CeuI, and the C-terminal subunit is derived from a single LAGLIDADG subunit from a wild-type di-LAGLIDADG meganuclease selected from I-DmoI, I-SceI and I-AniI.

**[0017]** Wild-type mono-LAGLIDADG meganucleases include, without limitation, the I-CreI meganuclease of SEQ ID NO: 1, the I-MsoI meganuclease of SEQ ID NO: 2, and the I-CeuI meganuclease of SEQ ID NO: 3. Wild-type di-LAGLIDADG meganucleases include, without limitation, the I-DmoI meganuclease of SEQ ID NO: 4, the I-SceI meganuclease of SEQ ID NO: 5, and the I-AniI meganuclease of SEQ ID NO: 6.

**[0018]** Wild-type LAGLIDADG domains include, without limitation, residues 9-151 of the wild-type I-CreI meganuclease of SEQ ID NO: 1; residues 11-162 of the wild-type I-MsoI meganuclease of SEQ ID NO: 2; and residues 55-210 of the wild-type I-CeuI meganuclease of SEQ ID NO: 3, residues 9-96 of the wild-type I-DmoI of SEQ ID NO: 4; residues 105-178 of the wild-type I-DmoI of SEQ ID NO: 4; residues 32-123 of the wild-type I-SceI of SEQ ID NO: 5; residues 134-225 of the wild-type I-SceI of SEQ ID NO: 5; residues 4-121 of the wild-type I-AniI of SEQ ID NO: 6; and residues 136-254 of the wild-type I-AniI of SEQ ID NO: 6.

**[0019]** 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-CreI meganuclease of SEQ ID NO: 1; residues 11-162 of the wild-type I-MsoI meganuclease of SEQ ID NO: 2; and residues 55-210 of the wild-type I-CeuI meganuclease of SEQ ID NO: 3, residues 9-96 of the wild-type I-DmoI of SEQ ID NO: 4; residues 105-178 of the wild-type I-DmoI of SEQ ID NO: 4; residues 32-123 of the wild-type I-SceI of SEQ ID NO: 5; residues 134-225 of the wild-type I-SceI of SEQ ID NO: 5; residues 4-121 of the wild-type I-AniI of SEQ ID NO: 6; and residues 136-254 of the wild-type I-AniI of SEQ ID NO: 6.

**[0020]** 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.

**[0021]** In certain embodiments, the invention provides 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.

**[0022]** In other embodiments, the invention provides 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.

**[0023]** In other embodiments, the invention provides 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.

**[0024]** In some embodiments, the polypeptide linker joining the LAGLIDADG subunits is a flexible linker. In particular embodiments, the linker can include 15-40 residues, 25-31 residues, or any number within those ranges. In other particular embodiments, at least 50%, or 50%-100%, of the residues forming the linker are polar uncharged residues.

**[0025]** In other embodiments, the polypeptide linker joining the LAGLIDADG subunits has a stable secondary structure. In particular embodiments, the stable secondary structure comprises at least two  $\alpha$ -helix structures. In other particular embodiments, the stable secondary structure comprises from N-terminus to C-terminus a first loop, a first  $\alpha$ -helix, a first turn, a second  $\alpha$ -helix, and a second loop. In some particular embodiments, the linker can include 23-56 residues, or any number within that range.

**[0026]** In another aspect, the invention provides for 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.

**[0027]** 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.

**[0028]** 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.

**[0029]** 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.

**[0030]** In another aspect, the invention provides methods of producing a 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 can be selected from a gamete, a zygote, a blastocyst cell, an embryonic stem cell, and a protoplast cell.

**[0031]** In another aspect, the invention provides 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.

**[0032]** Alternatively, in another aspect, the invention provides 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.

**[0033]** In another aspect, the invention provides 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.

**[0034]** In another aspect, the invention provides 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.

**[0035]** These and other 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

**[0036]** 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 linker.

## DETAILED DESCRIPTION OF THE INVENTION

### 1.1 Introduction

[0037] 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.

[0038] 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 homodimers. 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.

[0039] The invention also provides 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.

[0040] As a general matter, the invention provides 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 I-CreI, I-MsoI, or I-CeuI 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-SceI, I-DmoI, or I-AniI. 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.

[0041] 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.

[0042] Thus, in one aspect, the invention provides methods for producing recombinant single-chain LAGLIDADG meganucleases. In another aspect, the invention provides the recombinant single-chain meganucleases resulting from these methods. In another aspect, the invention provides 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, the invention provides 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

[0043] The patent and scientific literature referred to herein establishes knowledge that is available to those of skill in the art. The issued U.S. patents, allowed applications, published U.S. and PCT international applications, and references, including GenBank database sequences, that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference.

[0044] 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-SceI) or dimeric

(e.g., I-CreI). 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."

**[0045]** 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".

**[0046]** 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.

**[0047]** 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.

**[0048]** 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.

**[0049]** 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).

**[0050]** 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."

**[0051]** 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.

**[0052]** 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.

**[0053]** 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.

**[0054]** 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.

**[0055]** 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.

**[0056]** 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.

**[0057]** 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.

**[0058]** 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.

**[0059]** 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.

**[0060]** 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.

**[0061]** 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.

**[0062]** 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 frame-shift 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), *EMBO J.* 17:6086-6095). The process of non-homologous end-joining occurs in both eukaryotes and prokaryotes such as bacteria.

**[0063]** 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.

**[0064]** 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/](http://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.

**[0065]** 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.

[0066] 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  $\geq 0$  and  $\leq 2$  if the variable is inherently continuous.

[0067] 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

[0068] 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

**[0069]** 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**

<b>C-terminal amino acid</b>	<b>Wild-type activity</b>	<b>CCR2 activity</b>
Asp-153	+	+
Val-151	+	+
Val-148	+	-
Arg-141	-	-
Asn-136	-	-
Val-129	-	-
Ile-109	-	-
Leu-95	-	-

**[0070]** 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.

[0071] 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.

[0072] 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.

[0073] 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

[0074] 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).

[0075] 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.

[0076] 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)

Thus, the LAM1 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)

[0077] 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)

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

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

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

[0078] 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' -ACGCCACAGNNNNCTGTCTGGAC-5' (SEQ ID NO: 41)

### 2.2.1 Flexible linkers for single-chain meganucleases

[0079] 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).

**[0080]** 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)<sub>n</sub>, (GSS)<sub>n</sub>, (SGGS)<sub>n</sub>), may include random sequences, or may include combinations of the two.

**[0081]** 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-CreI 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) <sub>7</sub> G	-	-
2	Val-151	Phe-9	(GSS) <sub>8</sub> G	-	-
3	Val-151	Phe-9	(GSS) <sub>9</sub> G	+	+
4	Val-151	Phe-9	(GSS) <sub>10</sub> G	ND	+++
5	Val-151	Phe-9	(GSS) <sub>11</sub> G	ND	+++
6	Val-151	Phe-9	(GSS) <sub>9</sub> GG	+	+



7	Val-151	Phe-9	(GSS) <sub>9</sub> GSG	+	+
8	Asp-153	Phe-9	(GSS) <sub>9</sub> G	+	+

**[0082]** 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.

**[0083]** 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.

**[0084]** 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.

**[0085]** 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

**[0086]** 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](http://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):

[0087] (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  $\alpha$ -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- $\beta$ -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  $\alpha$ -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).

[0088] (2)  $\alpha$ -helix 1. Following Loop 1, this first  $\alpha$ -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 pphpphphpp 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  $\alpha$ -helical secondary structure. Alternatively, this helix sequence could be derived from a peptide sequence known to adopt  $\alpha$ -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](http://www.rcsb.org)).

[0089] In addition, it may be desirable to initiate the  $\alpha$ -helix with an N-capping motif to stabilize its structure (Aurora and Rose (1998), *Protein Sci.* 7:21-38). This motif spans the loop -  $\alpha$ -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).

**[0090]** (3) Turn 1. Following  $\alpha$ -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  $\alpha$ -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  $\alpha$ -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  $\alpha$ -helix 1 while the sequence PGSGI is Turn 1.

**[0091]** (4)  $\alpha$ -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.

**[0092]**  $\alpha$ -helix 2 can be designed to be shorter than  $\alpha$ -helix 1 and can comprise 1-3 turns of the helix (4-12 amino acids).  $\alpha$ -helix 2 can have the same overall amino acid composition as  $\alpha$ -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  $\alpha$ -helix 2 comprising the sequence SEALRA which is expected to make approximately two

turns. Experiments with different linker  $\alpha$ -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  $\alpha$ -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.

**[0093]** (5) Loop 2. This loop terminates  $\alpha$ -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  $\alpha$ -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  $\alpha$ -helix (amino acids 7-20) of the C-terminal I-CreI subunit. For example the sequence T-KSK<sub>7</sub>E<sub>8</sub>F<sub>9</sub> 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.

**[0094]** 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**  
**Linkers for Single-Chain I-CreI**

#	CFP	Linker Sequence	Activity	ppt
9	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGATKS	+++	-
10	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGGATKS	+++	-
11	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAAGGATKS	ND	++
12	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASGGATKS	ND	++
13	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASSGGATKS	ND	++
14	K7	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAASSAGGATKS	+++	-
15	E8	SLPGSVGGGLSPSQASSAASSASSSPGSGISEALRAGATKEF	++	+
16	E8	SLPGSVGGI <sub>1</sub> SPSQASSAASSASSSPGSGT <sub>1</sub> SEAP <sub>1</sub> RAGATKEF	++	-
17	E8	SLPGSVGGGLSPSQASSAASSASSSPGSGT <sub>1</sub> SEAT <sub>1</sub> RAGATKEF	++	+
18	E8	SLPGSLGGGLSPSQASSAASSASSSPGSGP <sub>1</sub> SEAL <sub>1</sub> RAGATKEF	++	+
19	E8	SLPGSVGGV <sub>1</sub> SPSQASSAASSASSSPGSGV <sub>1</sub> SEAS <sub>1</sub> RAGATKEF	++	+
20	E8	SLPGSVGGGLSPSQASSAASSASSSPGSGL <sub>1</sub> SEAL <sub>1</sub> RAGATKEF	++	+
21	E8	SLPGSLGGI <sub>1</sub> SPSQASSAASSASSSPGSGS <sub>1</sub> SEAS <sub>1</sub> RAGATKEF	++	-
22	E8	SPGSGVGGV <sub>1</sub> SPSQASSAASSASSSPGSGI <sub>1</sub> SEAT <sub>1</sub> RAGATKEF	++	-
23	E8	SLPGSLGGV <sub>1</sub> SPSQASSAASSP <sub>1</sub> SGT <sub>1</sub> SEAP <sub>1</sub> RAGATKEF	ND	++
24	E8	SLPGSVGGGLSPSQASSAASSP <sub>1</sub> SGI <sub>1</sub> SEAI <sub>1</sub> RAGATKEF	++	-
25	E8	SLPGSLGGV <sub>1</sub> SPSQASSAASSAASSP <sub>1</sub> SGA <sub>1</sub> SEAS <sub>1</sub> RAGATKEF	++	-

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).

[0095] 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-MsoI

[0096] I-MsoI 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-MsoI. Amino acids Phe-160, Leu-161, and Lys-162 of I-MsoI 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-MsoI. In addition, The X-ray crystal structure of I-MsoI 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-MsoI 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-MsoI (Table 7). In addition, because the sequence L<sub>7</sub>Q<sub>8</sub>P<sub>9</sub>T<sub>10</sub>E<sub>11</sub>A<sub>12</sub> of I-MsoI forms a natural N-cap (motif 2 from Table 5), Leu-7 can be included as a fusion point.



**TABLE 7**  
**I-MsoI Fusion Points**

<b>N-terminal fusion points</b>	<b>C-terminal fusion points</b>
Phe-160	Leu-7
Leu-161	Pro-9
Lys-162	Thr-10
Ile-166	Glu-11

**[0097]** Any of the linkers in Tables 3 or 6 can be used for the production of single-chain I-MsoI endonucleases. For example, Linker 9 from Table 6 may be used to join a pair of I-MsoI 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-CreI Lys-7) to Thr-10 (structurally homologous to I-CreI Glu-8).

**[0098]** 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,  $\alpha$ -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-MsoI**

#	NFP	CFP	Linker Sequence	Activity	ppt
26	K162	T10	PGSVGGLSPSQASSAASSASSSSPGSGISEALRAGATKSA	++	-
27	K162	T10	PGSVGGLSPSQASSAASSASSSSPGSGISEALRAGATKSG	++	-
28	I166	L7	GGASPSQASSAASSASSA <u>ASS</u> PGSGISEALRAASSLASKPGST	+++	-
29	I166	L7	GGASPSQASSAASSASSA <u>ASS</u> PGSGISEALRAASSPGST	+++	-
30	I166	L7	GGASPSQASSAASSASSA <u>ASS</u> PGSGPSEALRAASSFASKPGST	+++	-

#### 4. Single-chain meganucleases derived from I-CeuI

**[0099]** I-CeuI 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-CeuI. Amino acids Ala-210, Arg-211, and Asn-212 of I-CeuI 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-CeuI. Ser-53, Glu-54, and Ser-55 of I-CeuI 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-CeuI (Table 9).

**TABLE 9**  
**I-CeuI Fusion Points**

N-terminal fusion points	C-terminal fusion points
Ala-210	Ser-53
Arg-211	Glu-54
Asn-212	Ser-55

**[0100]** Any of the linkers in Tables 3 or 6 can be effective for the production of single-chain I-CeuI endonucleases. For example, I-CeuI 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.

**[0101]** The C-terminal fusion points selected for I-CeuI result in the removal of amino acids 1-52 from the C-terminal I-CeuI 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 of I-CeuI 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-CeuI 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- $\beta$ -branched, hydrophobic amino acids can be mutated to Ser while  $\beta$ -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.

**[0102]** 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)<sub>4</sub>G or (GSS)<sub>5</sub>G). 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<sub>212</sub>-SLPGSVGGLSPSQASSAASSASSSSPGS-G<sub>9</sub> C-term.  
subunit

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

**[0103]** 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.

**[0104]** 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**

<b>I-SceI</b> <b>N-terminal</b> <b>(31-123)</b>	<b>I-SceI</b> <b>C-terminal</b> <b>(132-225)</b>	<b>I-AniI</b> <b>N-terminal</b> <b>(3-125)</b>	<b>I-AniI</b> <b>C-terminal</b> <b>(135-254)</b>	<b>I-DmoI</b> <b>N-terminal</b> <b>(8-98)</b>	<b>I-DmoI</b> <b>C-terminal</b> <b>(104-178)</b>
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.

**[0105]** 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.

[0106] 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 LAGLIDADG 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.*

[0107] 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

[0108] 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

[0109] 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.

[0110] 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.

[0111] 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.

[0112] 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 be 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).

[0113] 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.

[0114] 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).

**[0115]** 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), *Adv. 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).

**[0116]** 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).

[0117] 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.

[0118] 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), *Proc. Natl. Acad. Sci. USA* 83:9065 9069.

[0119] 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  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989), *Genes Dev.* 3: 537-546).

[0120] 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

[0121] 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.

[0122] 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.

**[0123]** 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.

**[0124]** 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.

**[0125]** 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.

**[0126]** 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.

[0127] 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 deliver 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. Vasc. 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.

[0128] 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.

[0129] 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.

[0130] 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

[0131] 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.

[0132] 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  $\Delta F$ -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.

[0133] 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 D145 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**

[0134] This invention is further illustrated by the following examples, which should not be construed as limiting. Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific substances and procedures described herein. Such equivalents are intended to be encompassed in the scope of the claims that follow the examples below. 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.

### **EXAMPLE 1**

#### **Evaluation of the Method of Epinat *et al.***

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

[0135] 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.

[0136] 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.*

[0137] 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.

[0138] 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).

[0139] 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).

[0140] 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<sub>2</sub>, 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  $23,590 \text{ M}^{-1}\text{cm}^{-1}$ . Purity and molecular weight of the enzymes was then confirmed by MALDI-TOF mass spectrometry.

#### 4. Cleavage Assays.

[0141] 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  $\mu\text{l}$  0.05  $\mu\text{M}$  DNA substrate, 2.5  $\mu\text{l}$  5  $\mu\text{M}$  single-chain meganuclease, 9.5  $\mu\text{l}$  SA buffer, and 0.5  $\mu\text{l}$  *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

[0142] 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

[0143] 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 pET21a, expressed in *E. coli* and purified as described in Example 1.

#### 2. Results

[0144] 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-CreI Meganucleases Produced Using a Structured Linker

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

[0145] 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 pET21a, expressed in *E. coli* and purified as described in Example 1.

##### 2. Results

[0146] 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-MsoI Meganucleases Produced Using a Structured Linker

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

[0147] A pair of I-MsoI 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

[0148] Purified MSOdesMSO was assayed for the ability to cleave a plasmid substrate harboring the wild-type I-MsoI 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-MsoI 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-MsoI 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-MsoI.

**TABLE 11**  
**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
-1	Y75 L75* C75* Y139* C46* A46*	<b>R70*</b> H75* R75* H46* K46* R46*	K70 E70* E75* E46* D46*	Q70* C70 L70 Y75* Q75* H75* H139 Q46* H46*				<b>T46*</b>			G70 A70 S70 G46*
-2	Q70 T44* A44* V44* I44* L44* N44*	E70 D70 K44* R44*	H70 D44* E44*	<b>Q44*</b>	C44*						
-3	Q68 C24* <b>I24*</b>	E68 F68 K24* R24*	<b>R68</b>	M68 C68 L68 F68		H68		Y68	K68		
-4	A26* Q77	E77 K26*	R77 E26*					S77 <b>Q26*</b>			S26*
-5		E42	R42			<b>K28*</b>	C28* Q42				M66 K66
-6	Q40 C28*	E40 R28*	R40	C40 I40 V40 C79 I79 V79 Q28*	A40 A79 A28* H28*						<b>S40</b> S28*
-7	<b>N30*</b> <b>Q38</b>	E38 K30* R30*	K38 R38 E30*	I38 L38			C38				H38 N38 Q30*

**TABLE 11 (Continued)**

Posn. -8	Favored Sense-Strand Base										
	A F33 Y33	Posn. E33 D33	A F33 H33	Posn. L33 V33 I33 F33 C33	A	Posn. R32*	A R33	Posn.	A	Posn.	A
-9		E32	R32 K32	L32 V32 A32 C32				D32 I32			<b>S32</b> N32 H32 Q32 T32

Bold entries are 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 12****I-MsoI Modifications from WO 2007/047859**

Position	Favored Sense-Strand Base			
	A	C	G	T
<b>-1</b>	K75* Q77 A49* C49* K79*	<b>D77</b> E77 K49* <b>R75*</b> K75* R79* K79*	K77 R77 E49* E79*	C77 L77 Q79*
<b>-2</b>	Q75 K81 C47* I47* L47*	E75 D75 R47* K47* K81* R81*	K75 E47* E81*	A75 C75 V75 I75 T75 Q47* Q81*
<b>-3</b>	Q72 C26* L26* V26* A26* I26*	E72 Y72 H26* K26* R26*	<b>R72</b> K72 <b>Y26*</b> F26*	K72 Y72 H26*

TABLE 12 (Continued)

Position	Favored Sense-Strand Base			
	A	C	G	T
<b>-4</b>	<b>K28</b> Q83	<b>K28*</b> R28* E83	R83 K83	<b>K28</b> K83 Q28*
<b>-5</b>	<b>K28</b> C28* L28* I28*	<b>K28*</b> R28*	R45 E28*	Q28*
<b>-6</b>	<b>I30*</b> V30* S30* L30* Q43	E43 E85 K30* R30*	R43 K43 K85 R85 E30* D30*	K43 <b>I85</b> V85 L85 Q30*
<b>-7</b>	<b>Q41</b>	E32 E41	<b>R32</b> R41 K41	K32 M41 L41 I41
<b>-8</b>	<b>Y35</b> K35	E32	<b>R32</b> K32 K35 R35	K32 K35
<b>-9</b>	N34 H34	<b>D34</b> E34 S34	K34 R34 H34	S34 C34 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**

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

**TABLE 13 (Continued)**

<b>Position</b> <b>-9</b>	<b>Favored Sense-Strand Base</b>			
	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
	Q78	D78	R78	K78
	N78	E78	K78	V78
	H78		H78	L78
	K78			C78
				T78

Bold entries are 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 14****I-SceI Modifications from WO 2007/047859**

<b>Position</b>	<b>Favored Sense-Strand Base</b>			
	<b>A</b>	<b>C</b>	<b>G</b>	<b>T</b>
<b>4</b>	K50	<b>R50*</b> K50* E57	E50* R57 K57	K57 M57 Q50*
<b>5</b>	K48 <b>Q102</b>	<b>R48*</b> K48* E102 E59	E48* K102 R102	Q48* C102 L102 V102
<b>6</b>	K59	R59* K59*	K84 E59*	<b>Q59*</b> Y46
<b>7</b>	C46* L46* V46*	R46* K46* E86	<b>K86</b> R86 E46*	K68 C86 L86 Q46*
<b>8</b>	K61* S61* V61* A61* L61*	E88 R61* H61*	<b>E61*</b> <b>R88</b> K88	K88 Q61* H61*
<b>9</b>	<b>T98*</b> C98* V98* L98*	R98* K98*	E98* D98*	Q98*
<b>10</b>	<b>V96*</b> C96* A96*	K96* R96*	D96* E96*	Q96*

TABLE 14 (Continued)

Position 11	Favored Sense-Strand Base			
	A C90* L90*	C K90* R90*	G E90*	T Q90*
12	Q193	E165 E193 D193	K165 R165	C165 L165 C193 V193 A193 T193 S193
13	C193* L193*	<b>K193*</b> R193* D192	E193* D193* K163 R192	Q193* C163 L163
14	L192* C192*	E161 R192* K192*	K147 K161 R161 R197 D192* E192*	K161 Q192*
15		E151	K151	C151 L151 K151
17	<b>N152*</b> S152* C150* L150* V150* T150*	K152* K150*	N152* S152* D152* <b>D150*</b> E150*	Q152* Q150*
18	K155* C155*	R155* K155*	E155*	H155* Y155*

Bold entries are 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.

## CLAIMS

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. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of claim 1 wherein:
  - the first LAGLIDADG subunit is derived from a different species than the second LAGLIDADG subunit.
4. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of claim 2 wherein:  
each of said LAGLIDADG subunits has a recognition half-site selected from the group consisting of SEQ ID NOs: 7-30.
9. The recombinant single-chain meganuclease of claim 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. The recombinant single-chain meganuclease of any one of claims 1-9 wherein:  
said polypeptide linker is a flexible linker.
11. The recombinant single-chain meganuclease of claim 10 wherein:  
said linker comprises 15-40 residues.
12. The recombinant single-chain meganuclease of claim 10 wherein:  
said linker comprises 25-31 residues.
13. The recombinant single-chain meganuclease of claim 10 wherein:  
at least 50% of said linker comprises polar uncharged residues.

14. The recombinant single-chain meganuclease of any one of claims 1-9 wherein:  
said polypeptide linker has a stable secondary structure.
15. The recombinant single-chain meganuclease of claim 14 wherein:  
said stable secondary structure comprises at least two  $\alpha$ -helix structures.
16. The recombinant single-chain meganuclease of claim 14 wherein:  
said stable secondary structure comprises from N-terminus to C-terminus a first loop, a first  $\alpha$ -helix, a first turn, a second  $\alpha$ -helix, and a second loop.
17. The recombinant single-chain meganuclease of claim 14 wherein:  
said linker comprises 23-56 residues.
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
  - (i) a first nucleic acid sequence encoding a meganuclease, and
  - (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 claims 1-17.
19. A method as in claim 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. A method as in claim 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. 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 claims 1-17.

22. A method as in claim 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. A method as in claim 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. 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 claims 1-17.

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

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



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

26. A method as in claim 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. 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

- (i) a first nucleic acid sequence encoding a meganuclease, and
- (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 claims 1-17; and

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

28. A method as in claim 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. A method as in claim 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. 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 claims 1-17; and

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

31. A method as in claim 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. A method as in claim 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. 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 claims 1-17; and

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

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 claims 1-17; and

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

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 claims 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. 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 claims 1-17; and

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

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 claims 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.

1/1

**Figure 1****Structural Components of Linker 9**