



(51) International Patent Classification:

C12N 9/22 (2006.01) *C12Q* 1/34 (2006.01)
C12N 15/55 (2006.01) *C12Q* 1/68 (2006.01)

(74) Agents: **POOR, Brian, W.** et al.; Christensen O' Connor
 Johnson Kindness PLLC, 1420 Fifth Avenue, Suite 2800,
 Seattle, WA 98101 (US).

(21) International Application Number:

PCT/US2011/039527

(22) International Filing Date:

7 June 2011 (07.06.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/352,319 7 June 2010 (07.06.2010) US

(71) Applicants (for all designated States except US): **FRED HUTCHINSON CANCER RESEARCH CENTER** [US/US]; 1100 Fairview Avenue North, Seattle, WA 98109 (US). **SEATTLE CHILDREN'S HOSPITAL RESEARCH INSTITUTE** [US/US]; 1900 9th Avenue, Seattle, WA 98101 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **STODDARD, Barry, L.** [US/US]; 5260 116th Place S.E., Bellevue, WA 98006 (US). **LAMBERT, Abigail, Rose** [US/US]; 16614 Juanita Drive N.E., #203a, Kenmore, WA 98028 (US). **TAKEUCHI, Ryo** [JP/US]; 1903 N.E. 85th Street, #203, Seattle, WA 98115 (US). **SCHARENBERG, Andrew M.** [US/US]; 1222 NW Norcross Way, Seattle, WA 98177 (US). **BAXTER, Sarah Katherine** [US/US]; 4050 NE 57th Street, Seattle, WA 98105 (US).

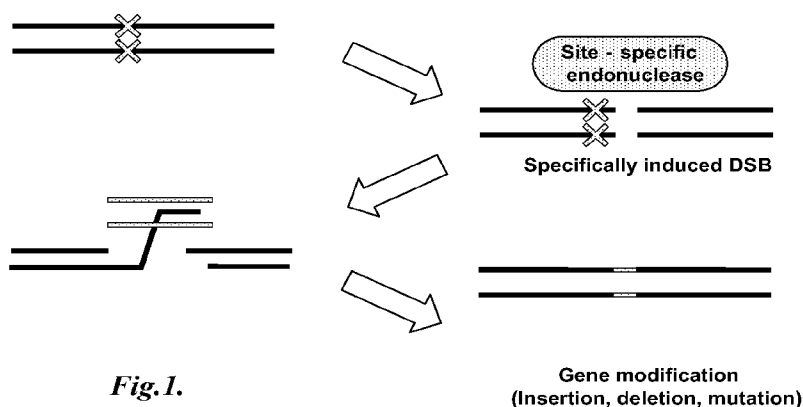
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, 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, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, 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 (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: GENERATION AND EXPRESSION OF ENGINEERED I-ONUI ENDONUCLEASE AND ITS HOMOLOGUES AND USES THEREOF



(57) Abstract: The present disclosure provides compositions and methods for producing and expressing variant or engineered I-Onul endonucleases, variant or engineered I-Onul homologues, and hybrids of two I-Onul or I-Onul homologue domains that have a target site altered from the wild-type. A method for selecting a variant or engineered I-Onul endonuclease, I-Onul endonuclease homologue, and a hybrid of two I-Onul or I-Onul homologue domains that have a target site altered from the wild-type and directed to a site within a gene of interest is also provided. In addition, the present disclosure provides the crystal structure of the I-Onul and I-LtrI endonucleases; the specificity profiles for both endonuclease for DNA binding and cleavage; the identity of amino acid residue positions in the I-Onul and I-LtrI protein scaffold that determine DNA recognition specificity; methods for determining amino acid substitutions at those positions that alter DNA cleavage specificity; methods for the complete redesign of the DNA cleavage specificity of I-Onul and its homologues for recognition and cleavage of a human gene of interest; and the relationship of the amino acid sequence, structure and specificity of I-Onul to a collection of identifiable I-Onul endonuclease homologues.

GENERATION AND EXPRESSION OF ENGINEERED I-ONUI ENDONUCLEASE
AND ITS HOMOLOGUES AND USES THEREOF

CROSS- REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application No. 61/352,319, filed
5 June 7, 2010, the disclosure of which is incorporated herein by reference.

STATEMENT REGARDING SEQUENCE LISTING

The sequence listing associated with this application is provided in text format in lieu
of a paper copy and is hereby incorporated by reference into the specification. The name of
the text file containing the sequence listing is: 37147_SEQ_FINAL.txt. The file is 153KB;
10 was created on June 7, 2011; and is being submitted via EFS-Web with the filing of the
specification

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY
SPONSORED RESEARCH AND DEVELOPMENT

The present invention was developed in part with government support under grant
15 numbers R01 GM49857, RL1 CA133833, RL1 CA133832 and UL1 DE019582 awarded by
the National Institutes of Health. The Government has certain rights in this invention.

BACKGROUND

The term "Genome Engineering" describes an emerging discipline in which genomes
of target organisms or cells are manipulated *in vivo*, generally using site specific
20 recombination or other modifications to alter or add genetic information at specific
chromosomal loci (through the targeted insertion, modification, or deletion of coding
sequence). The concept of genome engineering dates back to experiments in the late 1970s
in which ectopic DNA could be incorporated into the genome of the budding yeast
Saccharomyces cerevisia. (Hinnen *et al.*, *Proc. Nat'l. Acad. Sci. USA* 75:1929-1933, 1978;
25 Orr-Weaver *et al.*, *Proc. Nat'l. Acad. Sci. USA* 78:6354-6358, 1981). Depending on the

exact methodology, individual yeast genes could be efficiently incorporated, deleted, mutated or corrected. However, while homologous recombination is extremely efficient in yeast, in mammalian cells it occurs at a very low frequency, often in the range of 10^{-5} to 10^{-7} per transformed cells. (Doetschman *et al.*, *Nature* 330:576-578, 1987; Koller and Smithies, *Proc. Nat'l. Acad. Sci. USA* 86:8932-8935, 1989). As described below, this limitation can be in part overcome by using a highly site-specific endonuclease to cleave the donor or recipient locus to stimulate targeted recombination at a single chromosomal target site. The development of these reagents has allowed the field of genome engineering to progress dramatically over the past five years together with the pursuit of several specific applications. (Porteus, *Mol. Ther.* 13:438-446, 2006; Paques and Duchateau, *Curr. Gene Ther.* 7:49-66, 2007; Grizot *et al.*, *Nucl. Acids Res.* 37:5405-5419, 2009).

The use of a site-specific endonuclease or a recombinase to embed synthetic genes at specific desired target sites in model organisms represents a crucial enabling technology for synthetic biologists to create, manipulate, and control artificial genomes. (Collins *et al.*, *Nucl. Acids Res.* 38:2513, 2010). In particular, the ability to manipulate plant genomes in a controlled manner using targeted recombination is of great importance for the development of agricultural crop species both for food and for biofuel applications. (Porteus, *Nature* 459:337-338, 2009).

Synthetic genes encoding artificial site specific endonucleases can be used to create "selfish" genetic elements with the ability to integrate into and alter target genes while promoting their own transmission. This strategy has been proposed as a novel means for genetic control of *Anopheles*-mediated malaria transmission by dominant transmission and inheritance of traits corresponding to resistance against *Plasmodium* infection, or by reducing the lifespan or reproductive fitness of the insect host. (Chase, *Plant Sci.* 11:7-9, 2006).

Curative treatments for genetic diseases remain a critical area for research. (Ratjen and Dorign, *Lancet* 361:681-689, 2003; Griesenbach *et al.*, *Gene Ther.* 13:1061-1067, 2006).

Gene therapy approaches that rely on "gene augmentation" (*i.e.*, 'traditional' gene therapy, where a wild-type gene is integrated into a patient's somatic genome) are under active investigation. Many early problems associated with this area (poor gene delivery, immune reactions to viral delivery vehicles, and oncogenesis) are being addressed. (Griesenbach
5 *et al.*, *Gene Ther.* 13:1061-1067, 2006; Verma and Weitzman, *Annu. Rev. Biochem.* 74:711-738, 2005). However, the current practice of gene replacement therapy still has several attendant issues: First, gene therapy involves the random insertion of foreign DNA into the genomes of stem cells, potentially resulting in the inactivation or activation of endogenous genes. (Abbott, *Nat. Med.* 12:597, 2006; Hacien-Bey-Abina *et al.*,
10 *Science* 302:415-419, 2003; Themis *et al.*, *Mol. Ther.* 12:763-771, 2005). New lentiviral vectors avoid the use of highly active LTR-based promoters, and thus may improve safety profiles. (Griesenbach *et al.*, *Gene Ther.* 13:1061-1067, 2006).

A second issue for present gene replacement therapies is that it is desirable to use lineage-specific transcriptional control elements. However, defining such control elements is
15 non-trivial, and may require years of experimentation. (Puthenveetil *et al.*, *Blood* 104:3445-3453, 2004; Malik and Arumugam, *Hematology Am. Soc. Hematol. Edu. Program* 45-50, 2005; Malik *et al.*, *Ann. NY Acad. Sci.* 1054:238-249, 2005). A third issue is that traditional gene therapy is poorly suited for diseases caused by the presence of an aberrantly functioning protein that may interfere with function of the normal 'replacement'
20 protein. Finally, the problem of maintaining long-term protein expression after treatment remains problematic due to epigenetic silencing after gene integration.

Therefore, targeted "gene repair" or "gene correction" strategies, using highly specific endonucleases to stimulate homologous recombination and endogenous gene repair at a desired genomic target site, have been proposed. (Porteus, *Mol. Ther.* 13:438-446, 2006;
25 Paques and Duchateau, *Curr. Gene Ther.* 7:49-66, 2007). While gene repair has the same goal as traditional gene therapy approaches - restoration of the expression of a normally

functioning protein - it has many advantages. Since the endogenous gene's function is restored, the protein is expressed under the control of its natural regulatory elements, thus eliminating potential problems with inappropriate or inadequate expression of a transgene or transgene silencing. By targeting the repair with high efficiency to a single mutant locus,
5 gene repair may also be able to dramatically reduce mutagenesis due to random insertions at other genomic locations.

Several different technologies have been developed to promote efficient targeted genetic modification. These include gene-targeted triplex forming oligonucleotides and hybrid RNA-DNA oligonucleotides (Kolb *et al.*, *Trends Biotechnol.* 23:399-406, 2005) and
10 the use of highly site-specific recombinases and transposases (Coates *et al.*, *Trends Biotechnol.* 23:407-419, 2005). Each of these approaches has limitations related to the range of sequences that can be targeted (*e.g.*, triplex-forming oligonucleotides), or the requirement for prior introduction of a target site (*e.g.*, for recombinase-mediated targeting).

Potentially the most versatile of all genome engineering technologies are those that
15 make use of DNA double strand break-targeted homologous recombination for gene modification (FIGURE 1). This method allows a desired genomic sequence to be altered in a precise manner, without the requirement for a selection marker or the introduction of additional exogenous DNA sequence(s). Double strand break-targeted recombination requires the introduction or expression of a site-specific endonuclease in cells to generate a
20 DNA double strand break at or near the desired modification site, together with the presence of a DNA repair template. Repair templates typically flank the DNA double strand break site and include sequence modifications to be incorporated upon repair.

A significant practical barrier to widespread application of this accurate and efficient gene repair mechanism in genome engineering has been the requirement for an endonuclease
25 that is able to induce DNA double strand breaks at specific chromosomal target sites. Over the past several years, two different approaches to creating enzymes capable of inducing

highly site-specific DNA double strand breaks have been developed: zinc finger nucleases (ZFNs) and homing endonucleases (HEs).

A zinc finger nuclease is comprised of a non-specific nuclease domain (such as the catalytic domain of the *FokI* restriction endonuclease) tethered to a DNA-recognition and binding construct consisting of a tandem array of zinc fingers. (Porteus, *Mol. Ther.* 13:438-446, 2006; Smith *et al.*, *Nucl. Acids Res.* 28:3361-3369, 2000; Bibikova *et al.*, *Mol. Cell. Biol.* 21:286-297, 2001). As individual zinc fingers recognize DNA triplets within the context of long cognate target sites (Beerli and Barbas, *Nature Biotechnol.* 20:130-141, 2002; Bulyk *et al.*, *Proc. Nat'l. Acad. Sci. USA* 98:7158-7163, 2001; Segal *et al.*, *Proc. Nat'l. Acad. Sci. USA* 96:2758-2763, 1999), the concatenation of a series of zinc fingers of defined triplet specificity provides the possibility to create ZFNs able to bind and cleave at rare DNA targets. ZFNs have been demonstrated to induce gene correction in both *Drosophila* and mammalian cells (Bibikova *et al.*, *Science* 300:764, 2003; Porteus and Baltimore, *Science* 300:763, 2003), and the highly efficient correction of disease-associated mutations in the human IL2R γ gene (Urnov *et al.*, *Nature* 435:646-651, 2005). Zinc finger nucleases have the important advantage of some capacity for modular design, and therefore ZFN technology has been the subject of intensive study over the past ten years. (Porteus, *Mol. Ther.* 13:438-446, 2006).

Homing is the process by which mobile microbial intervening genetic sequences - group I or group II introns or inteins - are duplicated into host genes that lack such a sequence. (Dujon, *Gene* 82:91-114, 1989; Lambowitz and Belfort, *Annu. Rev. Biochem.* 62:587-622, 1993; Belfort and Perlman, *J. Biol. Chem.* 270: 30237-30240, 1995; Belfort and Roberts, *Nucl. Acids Res.* 25:3379-3388, 1997; Chevalier and Stoddard, *Nucl. Acids Res.* 29:3757-3774, 2001). This process is induced by a site-specific homing endonuclease encoded by an open reading frame (ORF) that is harbored within the intervening sequence. (Jacquier and Dujon, *Cell* 41:383-394, 1985). The endonuclease

specifically recognizes a target sequence corresponding to the intron insertion site and generates a single- or double-strand break that is repaired by cellular machinery. If the intron-containing allele is used as a template for repair via homologous recombination, the intron and its resident endonuclease gene is duplicated into the target site and the homing cycle is completed. Transfer of mobile introns can be extremely efficient, leading to unidirectional gene conversion events in diploid genomes (Jacquier and Dujon, *Cell* 41:383-394, 1985), genetic competition in mixed phage infections (Goodrich-Blair and Shub, *Cell* 84:211-221, 1996), gene transfer between different subcellular compartments of unrelated organisms (Turmel *et al.*, *Mol. Biol. Evolution* 12:533-545, 1995), and rapid genetic spread (Cho *et al.*, *Proc. Nat'l. Acad. Sci. USA* 95:14244-14249, 1998).

Homing endonucleases are widespread and found within introns and inteins in all biological super-kingdoms. At least six homing enzyme families have been identified (FIGURE 2); each is associated with a unique host genome. The LAGLIDADG endonuclease (LHE) are found in archaea, fungi and algae, the His-Cys Box family is found in protists, the HNH, GIY-YIG and VSR-like endonucleases (all found primarily in bacteriophage) and the PD-(D/E)xK family are found in bacteria.

In order to promote precise intron transfer and avoid deleterious cleavage of their host genomes, homing endonucleases are highly sequence-specific. However, they exhibit sufficient site recognition flexibility to promote genetic mobility in the face of target site variation across diverging host strains. Homing endonucleases use a strategy in which variable numbers of contacts are made to individual base pairs across a long target site (providing overall high specificity combined with variable recognition fidelity across the site). The individual polymorphisms that are tolerated by the enzyme are strongly correlated with the conservation of the sequence of the host target site. For a LAGLIDADG homing endonuclease (LHE), the specificity of DNA recognition is at least 1 in 10⁹. (Chevalier *et al.*, *J. Mol. Biol.* 329:253-269, 2003). By altering the DNA cleavage specificity of homing

endonucleases in the laboratory, a wide variety of such enzymes can be generated for genome engineering applications. (Paques and Duchateau, *Curr. Gene Ther.* 7:49-66, 2007). Several strategies have been investigated including: a) domain shuffling and fusions wherein domains from unrelated free-standing LHEs can be structurally fused to create chimeric

5 homing endonucleases that recognize corresponding chimeric target sites. (Chevalier *et al.*, *Molec. Cell.* 10:895-905, 2002). In addition, monomeric endonucleases can also be created from homodimeric proteins. (Li *et al.*, *Nucl. Acids Res.* 37:1650-1662, 2009); b) base-pair specificity changes using selections, screens and redesigns which include several methods that focus on mutation of endonuclease side chains that contact individual DNA basepairs to

10 alter specificity. These methods include: i) selections for efficient cleavage activity (Seligman *et al.*, *Nucl. Acids Res.* 30:3870-3879, 2002; Gruen *et al.*, *Nucl. Acids Res.* 30:e29, 2002, Sussman *et al.*, *J. Mol. Biol.* 342:31-41, 2004; Rosen *et al.*, *Nucl. Acids Res.* 34:4791-4800, 2006) or cleavage-induced homologous recombination events (Arnould *et al.*, *J. Mol. Biol.* 355:443-458, 2006; Chames *et al.*, *Nucl. Acids Res.* 33:e178, 2005);

15 ii.) structure-based computational redesign of DNA-contact surfaces and residues, to alter homing endonuclease-DNA contacts (Ashworth *et al.*, *Nature* 441:656-659, 2006; Ashworth *et al.*, *Nucl. Acids Res.* 38:5601-5608, 2010) or to facilitate efficient mutational screening of enzyme libraries (Sussman *et al.*, *J. Mol. Biol.* 342:31-41, 2004, Arnould *et al.*, *J. Mol. Biol.* 355:443-458, 2006; Chames *et al.*, *Nucl. Acids Res.* 33:e178, 2005); iii.) surface display

20 using either B-cells (Volna *et al.*, *Nucl. Acid Res.* 35:2748-2758, 2007) or yeast (Jarjour *et al.*, *Nucl. Acids Res.* 37:6871-6880, 2009) to facilitate characterization of binding and cleavage specificity profiles, and to sort populations of mutated endonucleases for shifts in binding specificity.

SUMMARY

25 The present disclosure provides compositions and methods for producing and expressing variant or engineered I-OnuI endonuclease, variant or engineered I-OnuI

homologues, and hybrids of two I-OnuI or I-OnuI homologue domains that have a target site altered from the wild-type. A method for selecting a variant or engineered I-OnuI endonuclease, I-OnuI homologue, and a hybrid of two I-OnuI or I-OnuI homologue domains with a target site modification from the wild-type and directed to a site within a gene of interest are also provided. The method for selecting a variant or engineered I-OnuI endonuclease comprises the steps of : i) determining the target site for an I-OnuI endonuclease; ii) searching a nucleic acid database for a gene of interest comprising a nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease; iii) selecting a gene of interest comprising the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease; iv) constructing a molecular model of the I-OnuI endonuclease bound to the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease from the gene of interest; v) mutating the I-OnuI endonuclease at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of variant or engineered I-OnuI endonuclease; vi) expressing the library of variant or engineered I-OnuI endonuclease; vii) screening the library of variant or engineered I-OnuI endonuclease for binding activity to the target sequence in the selected gene and the cleavage activity for the target sequence in the selected gene; and viii) selecting an altered, variant or engineered, I-OnuI endonuclease that can act upon a nucleotide sequence containing a modification in the target site from the wild-type and directed to the target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

The method for selecting a variant or engineered I-OnuI endonuclease homologue with a target site modification from the wild-type and directed to a site within a gene of interest comprises the steps of: i) determining the target site for a I-OnuI endonuclease

homologue; ii) searching a nucleic acid database for a gene of interest comprising a nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease homologue; iii) selecting a gene of interest comprising the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site
5 of the I-OnuI endonuclease homologue; iv) constructing a molecular model of the I-OnuI endonuclease homologue bound to the nucleotide sequence from the gene of interest that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease homologue; v) mutating the I-OnuI endonuclease homologue at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or
10 water-mediated contact residues with the target site of the gene of interest to form a library of variant or engineered I-OnuI endonuclease homologues; vi) expressing the library of variant or engineered I-OnuI endonuclease homologues; vii) screening the library of variant or engineered I-OnuI endonuclease homologues for binding activity to the target sequence in the selected gene and the cleavage activity for the target sequence in the selected gene; and
15 viii) selecting the variant or engineered I-OnuI endonuclease homologue that can act upon the modification in the target site from the wild-type and directed to the target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

The method for selecting an engineered hybrid of two I-OnuI or I-OnuI homologue
20 domains with a target site modification from the wild-type and directed to a site within a gene of interest comprises the steps of: i) determining the target site for a hybrid of two I-OnuI or I-OnuI homologue domains; ii) searching a nucleic acid database for a gene of interest comprising a nucleotide sequence that is at least 40% identical to the nucleotide sequence of target site of the hybrid of two I-OnuI or I-OnuI homologue domains;
25 iii) selecting a gene of interest comprising the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the hybrid of I-OnuI or I-OnuI

homologue domains; iv) constructing a molecular model of the hybrid of two I-OnuI or I-OnuI homologue domains bound to the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the hybrid of two I-OnuI or I-OnuI homologue domains from the gene of interest; v) mutating the hybrid of two I-OnuI or I-OnuI homologue domains at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of variant or engineered hybrids of two I-OnuI or I-OnuI homologue domains; vi) expressing the library of variant or engineered hybrids of I-OnuI or I-OnuI homologue domains; vii) screening the library of variant or engineered hybrids of I-OnuI or I-OnuI homologue domains for binding activity to the target sequence in the selected gene of interest and the cleavage activity for the target sequence in the selected gene of interest; and viii) selecting the engineered hybrid of two I-OnuI or I-OnuI homologue domains that can act upon the nucleotide sequence containing the modification in the target site from the wild-type and directed to a target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

Methods are also provided for producing an engineered endonuclease that can bind and cleave a specific site within a gene of interest. In this method a library of target sites for various endonuclease related to I-OnuI is established and the library is searched for a target site that has about 40% sequence identity with a nucleic acid sequence within the gene of interest. The selected endonuclease, homologue or hybrid can then be engineered. A method for producing an engineered I-OnuI endonuclease, an engineered I-OnuI endonuclease homologue, or a hybrid of two I-OnuI or I-OnuI homologue domains with a target site modification from the wild-type and directed to a site within a gene of interest comprises the steps of: i) determining the nucleotide sequence of the gene of interest; ii) searching a nucleic acid database comprising the target sites for I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains

for a I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains comprising a nucleotide sequence that is at least 40% identical to a nucleotide sequence of within the gene of interest; iii) selecting the I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains comprising the I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains with the target site that is at least 40% identical to the nucleotide sequence within the gene of interest; iv) constructing a molecular model of the selected I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains bound to the target site that is at least 40% identical to the nucleotide sequence within the gene of interest with the nucleic acid sequence within the gene of interest; v) mutating the selected I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of variants or engineered I-OnuI endonuclease, variant or engineered I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains; vi) expressing the library of variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains; vii) screening the library of variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains for binding activity to the target sequence in the gene of interest and the cleavage activity for the target sequence in the gene of interest; and viii) selecting the variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains that can act upon a nucleotide sequence containing a modification in the target site from the wild-type and directed to a target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

The present disclosure also describes variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues and variants thereof, and hybrids of two I-OnuI or I-OnuI homologue domain polypeptides, vectors comprising nucleic acid sequence that express variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues and variants thereof, and hybrids of two I-OnuI or I-OnuI homologue domain polypeptides, host cells, and various methods for the use of the variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologues and variants thereof, and hybrids of two I-OnuI or I-OnuI homologue domains.

DESCRIPTION OF THE DRAWINGS

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same become better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 depicts gene modification using a site-specific endonuclease. A double-strand break induced at or near a targeted chromosomal sequence ('X' on each strand) stimulates subsequent homologous recombination (HR), using an exogenous DNA template (typically provided by a plasmid transfection or a viral vector) as the donor of homologous DNA sequence. HR can result in targeted gene insertion, deletion, modification or mutation. In the absence of the donor DNA template, cleavage by the homing endonuclease can result in mutation of the target site as a result of nonconservative repair via DNA end-joining.

FIGURE 2 depicts the known structural families of homing endonucleases. Each protein fold shown above is named on the basis of conserved sequence motifs, and is largely restricted to individual biological hosts and corresponding genomes.

FIGURES 3A and B depict the expression and purification of I-OnuI. Figure 3A: The homing endonuclease was expressed as a fusion with an N-terminal

glutathione-S-transferase, purified by affinity chromatography and liberated by proteolytic processing with a protease designed for site specific cleavage of the GST tag (PreScission[®] protease; GE Healthcare). The protein was concentrated to 200 micromolar concentration in 50 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 20 mM MgCl₂, 5% glycerol. Figure 3B:

5 I-OnuI binds its naturally occurring target site (corresponding to the intron insertion site in the RP3 gene in its biological host with a dissociation constant (K_d) of approximately 37 picomolar, as determined by electrophoretic mobility shift analysis (EMSA: gel shift). The cleavage activity of the enzyme at a similar apparent enzyme concentration was confirmed using the same substrate sequence in a reporter plasmid (bottom).

10 FIGURE 4 provides the X-ray crystal structure of the I-OnuI homing endonuclease bound to its DNA target site. The structure was solved and refined at 2.4 angstrom resolution. The X-ray data and refinement statistics are provided in the table (right). The structure of the protein-DNA complex is displayed to the left in two separate orientations. The white circles highlight the 'LAGLIDADG' helices at the center of the interface between
15 the two endonuclease domains.

FIGURE 5 depicts the DNA-protein interface of I-OnuI. The 22 basepair target sequence (SEQ ID NO: 1 and SEQ ID NO: 2) is contacted by a combination of direct and water-mediated contacts, both to the DNA bases and its phosphoribosyl backbone. The scissile phosphates are white; the 'central four' basepairs between those phosphates are dark
20 grey. Waters are black spheres; bound divalent metal ions are light gray spheres.

FIGURES 6A and 6B provide the specificity profile of I-OnuI, determined using I-OnuI displayed on a yeast surface. Figure 6A demonstrates the relative ability of DNA sequences harboring individual basepair mismatches to be cleaved (the wild-type base at each position gives equivalent signal in the assay; bars with reduced heights indicate reduced
25 cleavage; bars with elevated heights (for example, -10A) indicate improved cleavage. The wild-type target sequence (SEQ ID NO: 1) is depicted at the top of the figure. Figure 6B

indicates the ability of the same target site variants to be recognized and bound by the enzyme. Those basepairs that can be bound and cleaved, in these separate experiments, as well or nearly as well as wild-type enzyme include -6C, -4C, +4T, and +4C.

FIGURE 7 depicts the designed and/or selected mutations in the I-OnuI enzyme scaffold that correspond to altered DNA cleavage specificity at individual basepair positions in the endonuclease target site. The positions and identity of the basepair substitutions are indicated in the left most column and by the base listed in bold in the target sites. SEQ ID NO: 3 through SEQ ID NO: 33. The corresponding mutations in the I-OnuI scaffold are indicated in the right column. Those that are shown in bold correspond to computationally designed DNA-contacting side chains (created using the crystal structure as a guide); those that are shown in italics correspond to mutations generated by selection experiments using an *in vivo* screen for cleavage and elimination of a reporter gene.

FIGURE 8 is an illustration of a human gene target (monoamine oxidase B) for targeted gene modification by a redesigned I-OnuI scaffold. The chromosomal locus and gene organization of MAO-B, a comparison of the nucleotide sequence of the target site (SEQ ID NO: 34) within MAO-B to be used for engineering a I-OnuI variant and the nucleotide sequence of the wild-type target site of I-OnuI (SEQ ID NO: 1) is also provided.

FIGURE 9 shows a summary of the selection of mutated variants of I-OnuI that display cleavage activity towards the MAO-B gene target (SEQ ID NO: 34). Mutation and selection of I-OnuI (SEQ ID NO: 35 through SEQ ID NO: 51) was conducted using an iterative approach in which individual basepair variants in the target site were incorporated in an iterative manner. Of the final "Round 3" constructs, R3 #3 (SEQ ID NO: 48), R3 #6 (SEQ ID NO: 50) and R3 #8 (SEQ ID NO: 51) displayed the highest cleavage activity towards the MAO-B target; one (R3 #3) was chosen for full characterization.

FIGURES 10A through 10C show the altered DNA cleavage specificity of the redesigned and selected "R3 #3" variant of I-OnuI. Figure 10A: titrations of wild-type (WT)

and redesigned (R3 #3) I-OnuI against WT target and MAO-B target. Figure 10B: Plots of cleavage progression against the WT and MAO-B DNA target sequences. Figure 10C: K_d and relative cleavage of wild-type and engineered variant of I-OnuI towards its target sites. Binding affinities were determined by electrophoretic gel shift analyses; k_{cat}/K_m by traditional endonuclease cleavage assays using radiolabeled oligonucleotide substrates. Non-cognate site cleavage by the wild-type and engineered versions of I-OnuI was undetectable; estimated detection limits for the assay are provided.

FIGURE 11 provides the protein sequence of the I-OnuI LAGLIDADG homing endonuclease and its immediate homologues (40% or higher sequence identity). All of these enzymes are encoded within algal and fungal organellar genomes, and are found in a wide variety of host genes (including those encoding several ribosomal proteins, cytochrome oxidases, ubiquinone oxidoreductase subunits, and ATP synthase subunits).

FIGURES 12A and 12B depict the crystal structure of I-LtrI and provides a schematic of its DNA binding surface. The crystal structure of I-LtrI bound to its DNA target sequence (SEQ ID NO: 86 and SEQ ID NO: 87) is shown in two different orientations and was solved to 2.7 Å resolution. Superposition of the structure on that of I-OnuI yields an RMSD across 284 superimposed α -carbons of approximately 1.3 Å and a similar DNA backbone configuration. The two -DNA interface of I-LtrI was quite dissimilar to that of I-OnuI. Only one side chain-nucleotide contact (between glutamine 195 and the adenine ring at base pair position +9) was observed in both structures. Taken together, these results suggest that even closely related LHEs such as I-OnuI and I-LtrI rapidly evolve unique, diverged surfaces for recognition of corresponding DNA target sites, while maintaining conserved protein folds and catalytic mechanisms.

FIGURES 13A through C depict the expression and cleavage activity of I-OnuI, I-LtrI and I-GpiI on yeast cell surface. Cells were stained with an anti-Myc probe (to visualize folding and surface expression; horizontal axis) and for binding of a DNA duplex containing

the validated I-OnuI (Figure 13A; SEQ ID NO: 88 and SEQ ID NO: 89) and I-LtrI (Figure 13B; SEQ ID NO: 90 and SEQ ID NO: 91) target sites or the predicted I-GpiI (Figure 13C; SEQ ID NO: 92 and SEQ ID NO: 93) DNA target site (vertical axis). Cleavage activity was visualized by loss of bound labeled DNA from the yeast cell surface in the presence of magnesium (which facilitates cleavage activity; dark gray population in each plot) relative to staining in the presence of calcium (which inhibited cleavage; lighter gray population in each plot).

Figure 14 demonstrates putative and validated DNA target sites for homologues of I-OnuI (see Figure 3). Target sites typically correspond to exon boundaries at the site of the mobile intron/homing endonuclease gene insertion in the host organism and genome (5' and 3' exons of the individual host genes are shown underlined and double underlined, respectively). Four endonucleases from this list (I-OnuI, I-LtrI, I-GpiI and I-MpeI) have been purified and tested against putative target sites, leading to validation of their DNA cleavage activities.

FIGURES 15A through D demonstrate the target sites and catalytic activity for a series of homologues of I-OnuI. Figure 15A depicts the target sites for I-OnuI (SEQ ID NO: 1), I-GpeI (SEQ ID NO: 94); I-LtrI (SEQ ID NO: 86); I-GpiI (SEQ ID NO: 95); I-MpeI (SEQ ID NO: 96); I-PanII (SEQ ID NO: 97); I-GzeI (SEQ ID NO: 98); I-SscI (SEQ ID NO: 99); I-AabI (SEQ ID NO: 100); I-PnoI (SEQ ID NO: 101); I-GzeII (SEQ ID NO: 102); I-CpaIII (SEQ ID NO: 103); I-LtrII (SEQ ID NO: 104); and I-SmaI (SEQ ID NO: 105). The cleavage of the predicted target sites were assayed using three different methods. Figure 15B depicts the results from a flow-cytometry based tethered DNA cleavage assay for I-GzeII. The flow cytometry-based assay used surface-expressed enzyme to cleave a fluorescently-tagged DNA substrate. Cleavage of the DNA sequence was visualized by a shift in the fluorescent signal. Figure 15C depicts the sequencing of the cleavage digest using a plasmid substrate. The target sequence for the enzyme I-GpiI (SEQ ID NO: 106 and SEQ ID NO:

107) was cloned into a circular plasmid and the plasmid substrate was digested by the enzyme. The resulting linearized DNA was sequenced from both sides of the break to determine the precise cleavage position on each strand of the DNA. This allowed for identification of the exact center of each target site. Figure 15D depicts an *in vitro* cleavage
5 digest for I-AabI. The surface-displayed enzyme was released from the yeast surface and used directly for an *in vitro* digest of the labeled DNA substrate. The cleaved oligonucleotide was distinguishable when visualized on an acrylamide gel.

FIGURE 16 depicts homology models of two I-OnuI homologues I-GpiI and I-MpeI. the models were based on the crystal structure of I-OnuI. Sequence alignments for these
10 endonucleases and others in the I-OnuI family are shown in Figure 11. Homology models of these and the other enzymes from Figure 11 have been used to engineer or select enzyme variants with altered surfaces and/or DNA contact residues for the purpose of altering their solution behavior, stability, and/or DNA binding and cleavage specificities (as illustrated in Figures 17 and 18)

15 Figures 17A through 17F depict the transfer of amino acid residues between the surfaces of I-OnuI-family homologues to alter protein solution behavior, folding, and/or DNA recognition properties. In initial studies were designed to transfer or "graft" all unconstrained surface from one homologue to another in order to achieve higher DNA sequence identity between individual enzymes, and to alter the protein's solution behavior
20 (Figures 17A through 17C). Approximately 40 to 60 mutations per homologue resulted in greater than 80% identity between the enzyme coding sequences, and excellent expression on the surface of yeast. DNA cleavage assays were performed to verify that activity was maintained throughout the engineering process. As an alternative, amino acid residues from the DNA-reaction surface can be grafted onto a different scaffold to successfully alter the
25 DNA target specificity of that scaffold (Figures 17D through 17F). This approach can be used to create artificial enzymes using DNA-contacting amino acid residues transplanted

between I-OnuI homologues. Figure 17A depicts a surface representation of I-OnuI with the location of solvent-exposed mutations highlighted in black. Figure 17B depicts the improvement in yeast surface expression of I-OnuI by incorporating surface amino acid mutations corresponding to polar amino acid residues transplanted from homologous homing endonucleases. The N- (APC) and C- (FITC) termini of the protein were fluorescently tagged to visualize stable full-length protein. Surface expression increased from 29% to 65%. Figure 17C depicts the verification of the cleavage activity of the "resurfaced" variant of I-OnuI (xOnu3) using an *in vitro* cleavage assay. Figure 17D provides a cartoon representation of the I-OnuI scaffold with positions of DNA-contacting residues and loops highlighted. These highlighted residues were transplanted from the I-MpeI endonuclease to the I-OnuI scaffold to create a different resurfaced variant termed "MpeItransOnuI". Figure 17E depicts the verification of stable surface expression for engineered MpeItransOnuI. Figure 17F depicts the verification of cleavage activity for MpeItransOnuI using the I-MpeI DNA target in an *in vitro* cleavage assay. The 'surface-transplanted' endonuclease now recognized and cleaved the target site originally recognized by the I-MpeI endonuclease.

Figure 18 shows the amino acid sequences of, and alignment between engineered variants of I-OnuI ("xOnuRound2")(SEQ ID NO: 108) and I-LtrI ("xLtrRound2")(SEQ ID NO: 109) that incorporate amino acid substitutions which increase sequence homology of each reading frame against one another (from 47% identity for the wild-type genes, to 76% for the redesigned endonucleases). The wild-type DNA cleavage specificities for these engineered variants were unchanged from wild-type scaffolds. This example demonstrates that hybrid, intermediate protein scaffold structures and sequences were fully accessible from the starting wild-type scaffolds in this protein family, and that further recombination and shuffling of these sequences can yield intermediate DNA cleavage specificities.

Figure 19 shows that hybrid homing endonucleases can be constructed from fusions of N- and C-terminal structural domains from I-OnuI and/or its identifiable homologues.

Chimeras of I-OnuI homologues are constructed by linking the N-terminal and C-terminal domains of two enzymes of interest (dark and light grey ribbon diagrams in the upper panel). Chimera assembly can be accomplished by gene synthesis, as well as by PCR gene-assembly of the respective halves, followed by restriction enzyme digestion and ligation of a shared site within the linker connecting the N- and C-terminal domains. The N-terminal domain is defined as the region approximately analogous to I-OnuI residues 1-162 (SEQ ID NO: 35), and the C-terminal domain is defined as the region approximately analogous to I-OnuI residues 163-303 (SEQ ID NO: 35). The exact location of N- and C-terminal division is variable, but lies within the 23 residues preceding the second LAGLIDADG helical region. Artificial linking residues can be included for ease of chimera construction, and for chimera optimization. Highly active chimera variants have been created using domains from I-OnuI and/or I-OnuI homologues I-GpiI, I-GzeI, I-SscI, I-PanII, and I-LtrI. The catalytic activity of constructed chimeras can be increased by randomization of protein residues within and near the LAGLIDADG helices and selection by flow-cytometry cleavage assays.

DETAILED DESCRIPTION

Generally, the nomenclature used herein and many of the laboratory procedures in regard to cell culture, molecular genetics and nucleic acid chemistry, which are described below, are those well known and commonly employed in the art. (See generally Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d Ed., Cold Spring Harbor Laboratory Press, New York (2001), which is incorporated by reference herein). Standard techniques are used for recombinant nucleic acid methods, site directed mutagenesis, preparation of biological samples, preparation of cDNA fragments, PCR, molecular modeling, crystallography, and the like. Generally enzymatic reactions and any purification and separation steps using a commercially prepared product are performed according to the manufacturers' specifications.

Homing endonucleases (HEs) are highly site-specific endonucleases that induce homologous recombination or gene conversion *in vivo* by cleaving long (typically greater than 20 basepair) DNA target sites. Homing endonucleases are under development as tools for applications that require targeted genome modification, including insertion, deletion, or modification of genetic coding sequences. The first structures of homing endonucleases were reported in 1997. Since that time, representative structures from each of the known families of homing endonuclease have been determined, and corresponding details of their mechanisms of DNA recognition and cleavage have been elucidated. Using this information, the LAGLIDADG homing endonuclease family (LHEs), which are distributed throughout single cell algal and fungi and in archaea, display the highest overall DNA recognition specificity. These proteins possess one or two LAGLIDADG catalytic motifs per protein chain and function as homodimers or as monomers, respectively. In addition, the family has been identified as the most tractable for further modification by structure-based selection and/or engineering approaches. To date, only the I-CreI-derived variants of the LAGLIDADG family have been used to modify endogenous chromosomal targets. Generation of these engineered I-CreI endonuclease was achieved by extensive alteration of the wild-type enzyme's DNA recognition specificity, at up to two-thirds of the base pair positions in a desired target site. The successful redesign of the I-CreI endonuclease has led to the development of enzymes that recognize and act at genes associated with monogenic diseases, including the human RAG1 and XPC genes (Redondo *et al.*, *Nature* 456:107-113, 2008, Grizot *et al.*, *Nuc. Acids Res.* 37:5405-5419, 2009), and the targeted disruption of at least one plant gene (LIGUELESS in corn) (Gao *et al.*, *Plant J.* 61:176-187, 2010). These studies demonstrate the feasibility of using engineered homing endonucleases to promote efficient and target site-specific modification of chromosomal loci. While this particular enzyme has been exceptionally cooperative during the process of protein engineering, the

reliance upon that single scaffold for genome engineering has limited the number of gene targets that can be modified using LHEs.

Use of many currently known homing endonucleases for the purpose of targeted gene insertion or modification is limited both by inappropriate biophysical behavior (for example, requiring thermophilic temperatures for DNA cleavage) or insufficient physiological cleavage activity, and by limitations on the extent which their DNA specificity can be altered. For any given enzyme, approximately one-third of the basepairs of its target site are not amenable to specificity-shifting redesign efforts. As well, extensive alteration of DNA binding specificity is often accompanied by losses of activity or affinity, or broadening of overall specificity, relative to the parental wild-type enzyme.

Therefore, there exists a need for the discovery, characterization, and engineering of a large collection of LAGLIDADG homing endonuclease scaffolds that possess the following characteristics: (1) monomeric (single chain) structures, (2) activity at physiological (30° to 37°C) temperatures, (3) high solubility and stability at physiological pH and ionic strength, (4) high enough amino acid identity across a broad range of homologous protein scaffolds (at least about 40 to 50%) to allow the creation of chimeric, hybrid, shuffled and recombined enzymes with high specificity and activity, and (5) sufficient diversity in their DNA sequence recognition profiles to allow recognition and cleavage of a much wider range of genomic targets than is currently possible with an individual homing endonuclease such as I-CreI.

The present disclosure demonstrates that naturally occurring LHEs can be exploited to rapidly create novel genome editing enzymes. In particular, the present disclosure focused on a single LHE subfamily (or 'clade') that are all related to the I-OnuI homing endonuclease, provides a surprisingly diverse set of DNA target site sequences that can be cleaved by these LHEs, that were otherwise closely related to one another. The target sites of these enzymes can be predicted and validated by analysis of the exon flanking sequences

that surround the homing endonuclease genes in their natural host cells. The disclosure also provides the determined DNA-bound crystal structure of two representative enzymes from this enzyme subfamily in order to assess the conservation of their protein folds and DNA recognition mechanisms, and then created a variant of one of those enzymes in order to
5 cleave and disrupt a predetermined human gene, the human monoamine oxidase B (MAO-B) gene. The present disclosure also demonstrates that hybrid enzymes can be created that contain distinct regions of multiple homing endonucleases, corresponding either to the transplantation and exchange of surface-exposed residues between enzyme homologues, or by fusion of unrelated N- and C- terminal domains between enzyme homologues. This
10 demonstrates that systematic mining and characterization of sufficient numbers of naturally occurring LHE scaffolds can allow the full potential of these enzymes for routine gene targeting applications to be realized.

As such, disclosed herein are (a) the crystal structures of the I-OnuI and I-LtrI homing endonucleases; (b) the specificity profile of I-OnuI for DNA binding and cleavage;
15 (c) the identity of amino acid residue positions in the I-OnuI and I-LtrI protein scaffolds that determine DNA recognition specificity; (d) examples of amino acid substitutions at those positions that alter DNA cleavage specificity; (e) an example of the complete redesign of the DNA cleavage specificity of I-OnuI for recognition and cleavage of a human gene therapy target (located in the monoamine oxidase-B gene target); (f) the relationship of the sequence,
20 structure and specificity of I-OnuI to a collection of identifiable I-OnuI endonuclease homologues with a wide variety of known and predicted DNA specificities; (g) expression and characterization of identifiable homologues of I-OnuI; and (h) creation and expression of hybrid scaffolds containing combined sequence elements of individual wild-type homing endonuclease, at a level of homology appropriate for future recombination and shuffling
25 experiments.

The endonuclease I-OnuI has been characterized by known methods to be a monomer displaying the characteristics of an LAGLIDADG homing endonuclease. As the molecule displayed certain characteristics required for gene targeting and subsequent engineerability the binding and cleavage activity of the isolated protein was determined. In addition, the crystal structure when bound to its substrate was determined. Further, homologues of I-OnuI were determined and are considered embodiments of the present disclosure that can be modified to alter their nucleic acid target sequence.

Homologues of I-OnuI can be identified by methods well known in the art. For example, sequence homology searches using I-OnuI as a query can be performed using the NCBI BLAST server (Altschul *et al.*, *J. Mol. Biol.* 215:403-410, 1990), using the BLASTP protein-protein alignment algorithm, using the NCBI-curated "non-redundant" protein databases; this resource corresponds to all current GenBank, RefSeq Nucleotides, EMBL (European Molecular Biology Laboratories), DDBJ (Databank of Japan), PDB (Protein Databank) and metagenomic sequences. For these searches, the default parameters for the search algorithm (including scoring matrix, gap penalties and compositional adjustments) are systematically varied until all recognizable homologues (corresponding to at about 25% or greater sequence identity extending over at least 200 amino acids and including both 'LAGLIDADG' sequence motifs) are identified. The present disclosure has determined that in certain embodiments, homologues will display highly conserved (greater than 50% sequence identity) to the LAGLIDADG motifs (amino acid residues 12 to 24 of SEQ ID NO: 35 and amino acid residues 170 to 181 of the I-OnuI amino acid sequence in SEQ ID NO: 35. In addition, homologues of I-OnuI will demonstrate high amino acid conservation (greater than 50% sequence identity) within a "Loop" sequence adjacent to the first LAGLIDADG helix corresponding to amino acid residues 97 to 103 of SEQ ID NO: 35. Still further, homologous of I-OnuI will demonstrate an overall spacing between the end and beginning of the two LAGLIDADG motifs of between about 162 and 182 amino acid

residues (See Figure 11). Individual homologues can display lower sequence identity at one of the regions described above, but conservation of the key residues involved in catalysis in the two separate LAGLIDADG motifs (corresponding to E22 or D22 in SEQ ID NO: 35 (E63 or D63 in Figure 11) and E178 or D178 in SEQ ID NO: 35 (E246 or D246 in Figure 11)) in the best clustered alignment of I-OnuI homologues allows their identification.

Homologues of I-OnuI (SEQ ID NO: 35) identified to date using the above criteria include, for example and not limitation, I-AabI (SEQ ID NO: 52), I-AaeI (SEQ ID NO: 53), I-ApaI (SEQ ID NO: 54), I-CkaI (SEQ ID NO: 55), I-CpaI (SEQ ID NO: 56), I-CapIII (SEQ ID NO: 57), I-CapIV (SEQ ID NO: 58), I-CpaV (SEQ ID NO: 59), I-CraI (SEQ ID NO: 60), I-EjeI (SEQ ID NO: 61), I-GpeI (SEQ ID NO: 62), I-GpiI (SEQ ID NO: 63), I-GzeI (SEQ ID NO: 64), I-GzeII (SEQ ID NO: 65), I-GzeIII (SEQ ID NO: 66), I-HjeII (SEQ ID NO: 67), I-LtrI (SEQ ID NO: 68), I-LtrII (SEQ ID NO: 69), I-MpeI (SEQ ID NO: 70), I-MveI (SEQ ID NO: 71), I-NcrI (SEQ ID NO: 72), I-NcrII (SEQ ID NO: 73), I-OheI (SEQ ID NO: 74), I-OsoI (SEQ ID NO: 75), I-OsoII (SEQ ID NO: 76), I-OsoIII (SEQ ID NO: 77), I-OsiIV (SEQ ID NO: 78), I-PanI (SEQ ID NO: 79), I-PanII (SEQ ID NO: 80), I-PanIII (SEQ ID NO: 81), I-PnoI (SEQ ID NO: 82), I-ScuI (SEQ ID NO: 83), I-SmaI (SEQ ID NO: 85), or I-SscI (SEQ ID NO: 86) as shown in Figure 11. Additional homologues of I-OnuI can be identified as above. Newly determined microbial and metagenomic sequence databases are expected to contain additional homologues of I-OnuI that can also be identified as above and can be used as embodiments in the methods described herein.

Using a computational model program or algorithm the crystal structure of I-OnuI bound to its DNA target site was used to determine information about the conformation of the I-OnuI homing nuclease nucleic acid binding site, including the detection, identification and optimization of contact points between individual I-OnuI domains (which can be used to create hybrid endonucleases that contain novel pairings of those domains) or between I-OnuI domains and substrate nucleic acid molecules (which can be used to create novel protein

DNA contacts and associated altered DNA binding and cleavage specificities). Similar methods and information were obtained for I-LtrI and can be obtained for any I-OnuI homologue, using a homology model of that homologue. A "homologue model" is a three dimensional model of a protein/DNA complex that is based upon the crystallographic structure of I-OnuI bound to its DNA target site. A "contact point" refers to the point at which the protein domains of I-OnuI, or a homologue thereof, or the protein domains of I-OnuI, or a homologue thereof, and its nucleic acid substrate interact. Such contact points are formed as a result of specific binding between two protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and a nucleic acid substrate molecule. Other amino acids within the interface can also be modified to enhance or improve the interaction between protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and nucleic acid molecules. "Interface" refers to the amino acids between protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and a nucleic acid molecule that form contact points, as well as those amino acids that are adjacent to contact points and along the planar surface between protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and nucleic acid molecules.

More importantly, the algorithm or program will allow the identification of either potential contact points or residues that are not properly interacting with a nucleic acid target sequence or other residues between protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and nucleic acid molecules that are inhibiting or reducing the overall interaction. Thus, the program or algorithm can identify potential contact points between protein domains of I-OnuI, or a homologue thereof, or between protein domains I-OnuI, or a homologue thereof, and nucleic acid molecules and/or identifying amino acids along the interface that can be modified to improve the

interface (that is improve the interaction between protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and a nucleic acid target sequence).

Points at which the unmodified I-OnuI, or a homologue thereof, usually directly
5 contacts a specific nucleic acid sequence (which may be different than the target sequence), but which are not contact points with the target sequence, are characterized as "potential contact points." A potential contact point can refer to one or more amino acids that usually directly interact with a nucleic acid sequence but are not stably interacting with the target sequence because a strong or stable enough chemical bond cannot be formed between the
10 amino acid(s) and the sequence. As a result, there is no contact point because of improper or inadequate bonding. The inability to bind can be due to chemical constraints (incompatible reaction groups) or proximity issues (too far or too close together). A specific chemical constraint can involve amino acids that repel each other or attract each other because of chemical charges. A specific proximity issue is when there is steric hindrance between either
15 amino acids or between an amino acid and the target sequence, which precludes or interferes with proper chemical bonding. Alternatively, an amino acid(s) can be too far from the target sequence to create an interface. In such cases, there is a gap between the two, which can be reduced or eliminated to create a contact point.

I-OnuI, a I-OnuI homologue, or a hybrid of two I-OnuI or I-OnuI homologue
20 domains, can be modified through one or more amino acid changes, including rotameric changes, to create an actual contact point between the I-OnuI, or homologue thereof, protein's nucleic acid binding domain and the target sequence. The I-OnuI or homologue thereof can also be modified through one or more amino acid changes to improve the interface between the protein and the nucleic acid. Methods for making these amino acid
25 changes are well known to the skilled artisan and are not consider a part of the present disclosure.

An amino acid change is a modification that is a substitution, deletion, or addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more contiguous or non-contiguous amino acids. Therefore, the present disclosure provides for the identification of an amino acid change that creates or enhances a contact point or the interface between individual protein domains of I-OnuI, or a homologue thereof, or between protein domains of I-OnuI, or a homologue thereof, and nucleic acid molecules, which can further provide a design for a modified I-OnuI polypeptide, or a homologue thereof. Enhancing a contact point or the interface means that a point between individual protein domains or between protein domains and nucleic acid molecules is made more chemically favorable, which includes reducing entropy, increasing stability, and reducing any steric hindrance.

Amino acid changes to create 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more interfaces may be determined empirically or computationally, or both. A change that is determined computationally refers to the use of a computer program or algorithm to identify amino acid changes that would create a desired contact point or improve or enhance the interface. Such programs are well known in the art. In some embodiments, the change is identified based on other polypeptides that interact with a site of similar sequence. Parameters known to those of ordinary skill in the art can be employed to guide the program or algorithm, such as sequence alignments, three-dimensional structural alignments, calculations of molecular interaction energies, and docking scores based on molecular complementarity.

Amino acid side chains from the two structurally independent, anti-parallel β -sheets (one from each protein domain of I-OnuI, or a homologue thereof), can be used to contact nucleotide bases within the major groove, at positions flanking the central four base pairs (FIGURE 5). The 22 base pair target site (SEQ ID NO: 1 and SEQ ID NO: 2), as shown in the illustration, was determined to be recognized by a combination of at least 22 mediated contacts between amino acid side chains and individual DNA bases, at least 14 additional

water mediated contacts between amino acid side chains and individual DNA bases, and at least 30 contacts between protein and the DNA phosphoribose backbone (mostly water mediated). Further, at least 40 amino acid residues were found to be involved in direct or water-mediated contacts to the DNA target; these represent the first shell of amino acid residues that can be exploited for a redesign or selection to either improve the binding and enzyme activity of the I-OnuI endonuclease with its wild-type target sequence or to modify the binding site recognized and cleaved by the I-OnuI endonuclease. Similar methods can be used to either improve or alter the binding specificity and/or activity of a I-OnuI endonuclease homologue.

"Hybrid" homing endonuclease as used herein is a protein scaffold that contains distinct, recognizable elements of sequence or structure contributed by two or more separate wild-type (*i.e.*, naturally occurring) homing endonucleases. A hybrid homing endonuclease can comprise any of the following combinations of protein elements contributed by separate wild-type homing endonuclease scaffolds: a) a fusion of two separate homing endonuclease domains (such as that previously described for "E-DreI" (now termed "H-DreI" for "hybrid I-DmoI/I-CreI") (Chevalier *et al.*, *Mol. Cell* 10:895-905, 2002) and as illustrated in Figure 19; b) a substitution of a peptide linker sequence from one homing endonuclease to connect the individual domains of a different homing endonuclease ; c) a substitution of a DNA-contacting peptide or loop from one endonuclease into the comparable region of a different homing endonuclease; d) the substitution of surface-displayed residues from one homing endonuclease onto comparable positions of a different homing endonuclease as illustrated in FIGURE 17; e) a fusion of multiple peptide regions from two or more homing endonuclease regions, leading to a single active endonuclease scaffold, using processes of DNA shuffling, recombination and PCR assembly from wild-type homing endonuclease coding sequences.

The binding and cleavage specificity of I-OnuI has been determined and is presented herein. The I-OnuI endonuclease was expressed using an expression vector expression/host cell system as described in detail in the examples. Methods for the expression and isolation of any endonuclease including I-OnuI, or a homologue thereof, are well known in the art.

5 The binding site or target site for I-OnuI is set forth in FIGURES 5 and 6. The specificity profile of I-OnuI endonuclease was determined as described in Jarjour *et al.* (*Nucl. Acids Res.* 37:6871-6880, 2009, incorporated herein by reference in its entirety). The specificity profiles illustrated the ability of I-OnuI to bind and cleave a series of alternative DNA target sequences that each contain a single basepair mismatch at each of the 22 positions in the

10 DNA recognition site. The analysis also indicated that the majority of specificity of DNA recognition was accomplished during DNA binding, rather than at the chemical step of DNA hydrolysis (although there were several individual basepair substitutions that did not affect binding, but inhibited subsequent cleavage). Those basepair substitutions that can be bound and cleaved as well (or at least with an efficiency greater than at least 50% that of the

15 corresponding wild-type DNA base) include -6C, -4C, +4T and +4C. The overall specificity of the enzyme was therefore extremely high (at least 1 in 10^{10}).

In one embodiment of the disclosure engineered variants of I-OnuI that can cleave altered DNA target sites (containing individual basepair substitutions) were identified using a combination of structure-based design and genetic selection for cleavage activity. The

20 methods relied upon the identification of amino acid contact residues near each altered DNA basepair that was identified in the protein-DNA crystal structure and subsequent selection of surrounding amino acid substitutions (corresponding to a 'pocket' of protein side chains that surround the immediate contacting residue at each position in the protein-DNA interface). The analysis identified alternative amino acid identities that could form contacts to altered

25 basepairs at each position in the DNA target, while modeling a conservatively flexible nucleotide and protein backbone. Methods, programs and algorithms capable of making

these calculations and determinations are well known in the art. Subsequent creation of limited protein mutation libraries using known methods, and screening these libraries for DNA cleavage that leads to elimination of a bacterial reporter gene has produced active enzyme variants with desired specificities for many single base-pair substitutions in the wild-type target site of I-OnuI homing endonuclease. In particular, mutations of Serine (S) at position 40 to Glutamic acid (E) (S40E), Asparagine at position 32 to Arginine (R) (N32R), Glutamic acid (E) at position 42 to glutamine (Q) (E42Q), Lysine 80 at position 80 to Arginine (R) (L80R), Serine (S) at position 78 to Glutamine (Q) (S78Q), Threonine (T) at position 48 to Lysine (K) (T48K), Glycine (G) at position 73 to Glutamic acid (E) (G73E), Valine at position 238 to Arginine (R) (V238R), Valine at position 199 to Arginine (R) (V199R), and Isoleucine at position 186 to Glutamine (Q) (I186Q) were found to recognize variant DNA cleavage sites. (See FIGURE 7 and SEQ ID NO: 35). In addition, in certain embodiments a combination of mutations were found to result in a variant I-OnuI endonuclease that recognized a variant DNA cleavage site. These combination include, for example, wherein Arginine (R) at position 30 of I-OnuI (SEQ ID NO: 35) was replaced with Cysteine (C), Glutamic acid (E) at position 42 was replaced with Leucine (L), Threonine (T) at position 82 was replaced with Lysine (K), Arginine at position 83 was replaced by Valine (V), Leucine (L) at position 87 was replaced by Phenylalanine (F), and Isoleucine (I) at position 90 was replaced with Methionine (M) (R30C/E42L/T82K/R83V/L87F/I90M); wherein the Serine at position 72 was replaced by Alanine (A), Asparagine (N) at position 75 was replaced by Arginine (R), and Alanine (A) at position 76 was replaced by Leucine (L) (S72A/N75R/A76L); wherein the Serine (S) at position 201 was replaced by Glutamine (Q), Lysine (K) at position 227 was replaced by Glycine (G) and Aspartic acid (D) at position 236 was replaced by Valine (S201Q/K227G/D236V); wherein the Asparagine (R) at position 184 was replaced by Alanine, Valine (V) at position 199 was replaced by Arginine (R), Lysine (K) at position 225

was replaced by Histidine (H) (N184A/V199R/K225H); and wherein the Asparagine (N) at position 184 was replaced by Threonine (T), Isoleucine at position 186 was replaced by Glutamine (Q), Glutamine (Q) at position 197 was replaced by Arginine (R), and Valine at position 199 was replaced by Glutamic acid (E) (N184T/I186Q/Q197R/V199E). The selection system was also modified to allow for the selection of an enzyme specificity as well as activity by selecting against enzyme variants that could still cleave the wild-type target site.

One of the main purposes of the present disclosure is to obtain a specific endonuclease that can cleave a specific site within a gene of interest. As such, a variant of I-OnuI was engineered to recognize and cleave a physiological target site in the human genome. The human monoamine oxidase B gene was selected because the gene includes a nucleotide sequence highly similar to the target nucleotide sequence of I-OnuI. See FIGURE 8 and SEQ ID NO: 34. In addition, human monoamine oxidase B (MAO-B) catalyzes the deamination of a large number of biogenic amines in the brain and central nervous system, including serotonin, dopamine, and phenylethylamine (Shih *et al.*, *Annu. Rev. Neurosci.* 22:197-217, 1999). Point mutations in MAO-B (as well as in monoamine oxidase A) are associated with many neurological and cognitive disorders, including Parkinson's Disease, compulsive and addictive behaviors, stress disorders, and aggressive behaviors (Shih *et al.*, *Annu. Rev. Neurosci.* 22:197-217, 1999; Lew, *Pharmacotherapy* 27:155S-160S, 2007). Although this gene is not typically a gene therapy target, MAO-B is under intense study (primarily using gene knockouts in mice) to elucidate the role of wild-type and mutant enzymes in neurological function and behavior (Grimsby *et al.*, *Nature Genet.* 17:1-5, 1997). Using the crystal structure obtained for I-OnuI and its target DNA sequence and the models for the I-OnuI binding sites, variants of I-OnuI have been engineered that target a unique sequence in the MAO-B target site. This variant I-OnuI can be used to direct homologous recombination in that gene locus, leading to targeted point

mutations of the endogenous MAO-B gene and the development of neural cell lines for *in vivo* studies.

The ability to engineer a variant I-OnuI endonuclease to a new target site demonstrated the concept that nucleotide target sequences having at least 40% identity to the target sequence of I-OnuI, or a homologue thereof, in a gene of interest can be selected and that a variant of I-OnuI or a homologue thereof (including hybrid endonucleases) can be made that can efficiently bind to and cleave the DNA within the gene at this specific site. Further, if the nucleotide sequence of a gene of interest is known the binding sites for each of the I-OnuI family of endonucleases can be searched to find a binding site sufficiently related (at least 40% identity) that can be modified by the methods discloses herein to target a variant binding site within the gene of interest. As such, the altered and hybrid endonucleases as described herein make possible targeted gene insertion or modification in a greater number of genes of interest. The family of I-OnuI endonuclease and its homologues provided a broad set of protein scaffolds for the design and selection of additional DNA cleavage specificities.

As such, the method for selecting an engineered I-OnuI endonuclease, a I-OnuI endonuclease homologue, or a hybrid of two I-OnuI or I-OnuI homologue domains with a target site modification from the wild-type and directed to a site within a gene of interest comprises the steps of; i) determining the target site nucleic acid sequence for the I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains; ii) searching a nucleic acid database for a nucleotide sequence at least 40% identical to the target site nucleotide sequence of the I-OnuI endonuclease or I-OnuI homologue endonuclease, or a hybrid of two I-OnuI or I-OnuI homologue endonuclease domains; iii) selecting a gene of interest having the nucleotide sequence at least 40% identical to the target site nucleic acid sequence of the I-OnuI endonuclease, I-OnuI homologue endonuclease, or the hybrids of two I-OnuI or I-OnuI homologue endonuclease

domains; iv) mutating the I-OnuI endonuclease, I-OnuI homologue endonuclease, or the hybrid of two I-OnuI or I-OnuI homologue endonuclease domains at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the nucleic acid sequence of the gene of interest having
5 at least 40% identity to the target site of the I-OnuI endonuclease, I-OnuI homologue endonuclease, or the hybrid of two I-OnuI or I-OnuI homologue endonuclease domains to form a library of variant or engineered I-OnuI endonuclease, I-OnuI homologue endonuclease or variants thereof, or hybrids of two I-OnuI or I-OnuI homologue endonuclease domains; v) expressing the library of engineered I-OnuI endonuclease, I-OnuI
10 homologue endonuclease, or hybrids of two I-OnuI or I-OnuI homologue endonuclease domains; vi) screening the library of variant or engineered I-OnuI endonuclease, I-OnuI homologue endonuclease and variants thereof, or hybrids of two I-OnuI or I-OnuI homologue endonuclease domains for binding activity to the nucleic acid sequence in the selected gene of interest and the cleavage activity for the target sequence in the selected gene
15 of interest; and vii) selecting the variant or engineered I-OnuI endonuclease, I-OnuI endonuclease homologue or variant thereof, or the hybrid of two I-OnuI or I-OnuI homologue endonuclease domains with the highest binding and cleavage activity for the target sequence in the selected gene of interest.

In another embodiment, the method can be carried out starting from a selected gene
20 of interest and searching a data base of I-OnuI endonuclease and I-OnuI endonuclease homologue target sequences. Once the target sequences of the I-OnuI endonuclease or the I-OnuI endonuclease homologue, or the hybrid of two I-OnuI or I-OnuI homologue domains with the best match with a nucleotide sequence within the target gene has been selected steps
iv) through viii) above can be repeated and the variant or engineered I-OnuI endonuclease,
25 I-OnuI endonuclease homologue or variant thereof, or hybrid of two I-OnuI or I-OnuI homologue domains selected for the highest binding activity and/or cleavage activity for the

target sequence in the gene of interest. Once selected the variant or engineered I-OnuI endonuclease, I-OnuI homologue or variant thereof, and/or the engineered hybrid of two I-OnuI or I-OnuI homologue endonuclease domains can be used as set forth below. The term "engineered I-OnuI endonuclease, or homologues and hybrids thereof" or "variants of I-OnuI" as used herein include the engineered I-OnuI endonucleases, engineered I-OnuI homologues, and/or the engineered hybrid of two I-OnuI or I-OnuI homologue endonuclease domains set forth above.

Variants of I-OnuI, including variants of I-OnuI and I-OnuI homologues, *e.g.*, I-LtrI, and hybrids of two I-OnuI or I-OnuI homologue domains, that can cleave altered DNA target sites (containing individual basepair substitutions) were identified using a combination of structure-based design and genetic selection for cleavage activity. The method relied upon identification of amino acid contact residues near each altered DNA basepair that were identified in the protein-DNA crystal structure (such as, for example, illustrated in FIGURE 5 and Figure 12B) and subsequent selection of surrounding amino acid substitutions (corresponding to a 'pocket' of protein side chains that surround the immediate contacting residue at each position in the protein-DNA interface). The analysis can identify alternative amino acid identities that can form contacts to altered basepairs at each position in the DNA target, while modeling a conservatively flexible nucleotide and protein backbone. Subsequent creation of limited protein mutation libraries and screening these libraries for DNA cleavage that leads to elimination of a bacterial reporter gene has produced active enzyme variants with desired specificities for many single base-pair substitutions in the wild-type target site of I-OnuI homing endonuclease (See, for example, FIGURE 7). The selection system was also modified to allow for selection of enzyme specificity as well as activity by selecting against enzyme variants that can still cleave the wild-type target site. In one embodiment the engineering of a variant of I-OnuI against a physiological target site in the human genome (the monoamine oxidase B (MAO-B) gene) is described below.

Provided herein are also vectors comprising the nucleic acid sequence that encodes the variant engineered I-OnuI endonuclease, I-OnuI endonuclease homologue, or engineered hybrid of two I-OnuI and/or I-OnuI homologue endonuclease domains of the present disclosure. A nucleic acid encoding one or more engineered I-OnuI, or engineered I-OnuI homologue or hybrid thereof, can be cloned into a vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Vectors can be prokaryotic vectors, *e.g.*, plasmids, or shuttle vectors, insect vectors, or eukaryotic vectors. A nucleic acid encoding an engineered I-OnuI, or homologue or hybrid thereof, as disclosed herein can also be cloned into an expression vector, for administration to a plant cell, animal cell, a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoan cell.

To obtain expression of a cloned gene or nucleic acid, sequences encoding an engineered I-OnuI, or homologue or hybrid thereof, is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, *e.g.*, in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2nd ed. 1989; 3rd ed. , 2001); Kriegler, Gene Transfer and Expression. A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel *et al.*, *supra.*) Bacterial expression systems for expressing the engineered I-OnuI, and homologue or hybrid thereof, are available in, *e.g.*, *E. coli*, *Bacillus* sp., and *Salmonella* (Palva *et al.*, *Gene* 22:229- 235, 1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known by those of skill in the art and are also commercially available.

The promoter used to direct expression of an engineered I-OnuI, or homologue or hybrid thereof encoding nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of a variant or engineered I-OnuI, or homologue or hybrid thereof. In contrast, when an engineered I-OnuI, or homologue or hybrid thereof, is administered *in vivo* for gene regulation, either a

constitutive or an inducible promoter is used, depending on the particular use of the engineered I-OnuI, or homologue or hybrid thereof. In addition, a promoter for administration of an engineered I-OnuI, or homologue or hybrid thereof, can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically
5 can also include elements that are responsive to transactivation, *e.g.*, hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (see, *eg.*, Gossen and Bujard, *Proc. Nat'l. Acad. Sci.* 89:5547, 1992; Oligino *et al.*, *Gene Ther.* 5:491-496, 1998; Wang *et al.*, *Gene Ther.* 4:432-441, 1997; Neering *et al.*, *Blood* 88:1147-1155, 1996; and
10 Rendahl *et al.*, *Nat. Biotechnol.* 16:757-761, 1998). The MNDU3 promoter can also be used, and is preferentially active in CD34⁺ hematopoietic stem cells.

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in a host cell, either prokaryotic or eukaryotic. A typical
15 expression cassette thus contains a promoter operably linked, *e.g.*, to a nucleic acid sequence encoding the engineered endonuclease, and signals required, *e.g.*, for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, *e.g.*, enhancers, and heterologous splicing signals.

The particular expression vector used to transport the genetic information into the cell
20 is selected with regard to the intended use of the engineered I-OnuI, or homologue or hybrid thereof, *e.g.*, expression in plants, animals, bacteria, fungus, protozoa, and the like. Standard bacterial expression vectors include plasmids such as pBR322- based plasmids, pSKF, pET23D, pBluescript[®] based plasmids, and commercially available fusion expression
25 systems such as GST and LacZ. An exemplary fusion protein is the maltose binding protein, "MBP." Such fusion proteins are used for purification of the engineered I-OnuI, or

homologue or hybrid thereof. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, *e.g.*, a nuclear localization signal (NLS), an HA-tag, c-myc or FLAG.

5 Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A⁺, pMT010/A⁺, pMAMneo-5, baculovirus pDSVE, pCS, pEF, and any other
10 vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, elongation factor 1 promoter, or other promoters shown effective for expression in eukaryotic cells.

 Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase.
15 High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with an engineered I-OnuI, or homologues thereof, encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

 The elements that are typically included in an expression vector also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of
20 bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

 Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (Guide to Protein Purification, in Methods in Enzymology, vol. 182
25 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques well known to the skilled artisan.

Any of the well known procedures for introducing foreign nucleotide sequences into host cells can be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, ultrasonic methods (*e.g.*, sonoporation), liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, *e.g.*, Sambrook et al., *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

Nucleic acids encoding an engineered I-OnuI, or homologue or hybrid thereof, as described herein and delivery to cells can use conventional viral and non-viral based gene transfer methods (*e.g.*, mammalian cells) and target tissues. Such methods can also be used to administer nucleic acids encoding an engineered I-OnuI, or homologue or hybrid thereof, to a cell *in vitro*. In certain embodiments, nucleic acids encoding an engineered I-OnuI, or homologue or hybrid thereof, are administered for *in vivo* or *ex vivo* gene modification uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell.

Methods of non-viral delivery of nucleic acids encoding an engineered I-OnuI, and homologue or hybrid thereof, include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid: nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

Lipofection is described in *e.g.*, US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (*e.g.*, Transfectam and LipofectinTM). Cationic

and neutral lipids that are suitable for efficient receptor recognition of polynucleotides include those of Feigne, WO1991/17424, WO1991/16024. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration). The preparation of lipid: nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art.

The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding an engineered I-OnuI, or homologue or hybrid thereof, takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients or they can be used to treat cells *in vitro* and the modified cells are administered to patients. Conventional viral based systems for the delivery of an engineered I-OnuI, or homologue or hybrid thereof, include, but are not limited to, retroviral, lentivirus, adenoviral, adeno-associated, vaccinia and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system depends on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia

virus (GaLV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof.

In applications in which transient expression of an engineered I-OnuI, or homologue or hybrid thereof, is preferred, an adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and high levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene modification procedures. Construction of recombinant AAV vectors are described in a number of publications, including US 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260, 1985; Tratschin *et al.*, *Mol. Cell. Biol.* 4:2072-2081, 1984; Hermonat and Muzyczka, *Proc. Natl. Acad. Sci.* 81:6466-6470, 1984; and Samulski *et al.*, *J. Virol.* 63:03822-3828, 1989. Recombinant adeno-associated virus vectors (rAAV) provide an alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus.

Replication-deficient recombinant adenoviral vectors (Ad) can be produced at high titer and readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the AdE1a, E1b, and/or E3 genes; subsequently the replication defective vector is propagated in human 293 cells that supply deleted gene function in trans. Ad vectors can transduce multiple types of tissues *in vivo*, including nondividing, differentiated cells such as those found in liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity.

Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene modification are usually generated by a

producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host (if applicable), other viral sequences being replaced by an expression cassette encoding the protein to be expressed. The missing viral functions are typically supplied *in trans* by the packaging cell line. For example, AAV vectors used in gene modification typically only possess inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line can also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

Gene modification vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene modification (*e.g.*, via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a typical embodiment, cells are isolated from the subject organism, transfected with an engineered I-OnuI, or homologue or hybrid thereof, nucleic acid (gene or cDNA), and re-infused back into the subject organism (*e.g.*, a patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (see, *e.g.*, Freshney *et al.*,

Culture of Animal Cells, A Manual of Basic Technique (5th ed. 2005)) and the references cited therein for a discussion of how to isolate and culture cells from a patient).

Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, and the like) containing an engineered I-OnuI, or homologue or hybrid thereof, nucleic acid can also be administered directly to an organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available (see, *e.g.*, Remington The Science and Practice of Pharmacy, 21st ed., 2005).

DNA constructs may be introduced into the genome of a desired plant host by a variety of conventional techniques. For reviews of such techniques see, for example, Weissbach and Weissbach, Methods for Plant Molecular Biology (1988, Academic Press, N. Y.) Section VIII, pp. 421-463; and Grierson and Corey, Plant Molecular Biology (1988, 2d Ed.), Blackie, London, Ch. 7-9. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment.

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector.

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature and will not be described here.

Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-, polyethylene glycol (PEG)- or electroporation-mediated uptake of naked DNA and electroporation of plant tissues. Additional methods for plant cell transformation include microinjection, silicon carbide mediated DNA uptake, and microprojectile bombardment.

The disclosed methods and compositions can be used to make genomic changes and/or to insert exogenous sequences into a predetermined location in a plant cell genome. This is useful inasmuch as expression of an introduced transgene into a plant genome depends critically on its integration site. Accordingly, genes encoding, *e.g.*, nutrients, antibiotics or therapeutic molecules can be inserted, by targeted recombination, into regions of a plant genome favorable to their expression.

Transformed plant cells which are produced by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is a well known technique to the skilled artisan. Regeneration can also be obtained from plant callus, explants, organs, pollens, embryos or parts thereof.

Nucleic acids introduced into a plant cell can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems can be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned

above. Typically, target plants and plant cells for modification include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (*e.g.*, wheat, maize, rice, millet, barley, and the like), fruit crops (*e.g.*, tomato, apple, pear, strawberry, orange, and the like), forage crops (*e.g.*, alfalfa, and the like), root vegetable crops (*e.g.*, carrot, potato, sugar beets, yam, and the like), leafy vegetable crops (*e.g.*, lettuce, spinach, and the like); flowering plants (*e.g.*, petunia, rose, chrysanthemum, and the like), conifers and pine trees (*e.g.*, pine fir, spruce, and the like); plants used in phytoremediation (*e.g.*, heavy metal accumulating plants, and the like); oil crops (*e.g.*, sunflower, rape seed (canola), camelina, and the like) and plants used for experimental purposes (*e.g.*, Arabidopsis, and the like).

One of skill in the art will recognize that after the expression cassette is stably incorporated in a transgenic plant and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for traits encoded by the marker genes present on the transforming DNA. For instance, selection may be performed by growing the engineered plant material on media containing an inhibitory amount of the antibiotic or herbicide to which the transforming gene construct confers resistance. Further, transformed plants and plant cells may also be identified by screening for the activities of any visible marker genes (*e.g.*, the β -glucuronidase, green fluorescent protein, luciferase, B or Cl genes) that may be present on the recombinant nucleic acid constructs. Such selection and screening methodologies are well known to those skilled in the art.

Physical and biochemical methods also may be used to identify plant or plant cell transformants containing inserted gene constructs. These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of

the recombinant DNA insert; 2) Northern blot, siRNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, Western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the gene construct products are proteins. Additional techniques, such as *in situ* hybridization, enzyme staining, and immunostaining, also can be used to detect the presence or expression of the recombinant construct in specific plant organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

Effects of gene manipulation using an engineered I-OnuI endonuclease, or homologue or hybrid thereof, disclosed herein can be observed by, for example, northern blots of the RNA (*e.g.*, mRNA) isolated from the tissues of interest. Typically, if the amount of mRNA has increased, it can be assumed that the corresponding endogenous gene is being expressed at a greater rate than before. Other methods of measuring gene activity can be used.

Different types of enzymatic assays can be used, depending on the substrate used and the method of detecting the increase or decrease of a reaction product or by-product. In addition, the levels of and/or CYP74B protein expressed can be measured immunochemically, *i.e.*, ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art, such as by electrophoretic detection assays (either with staining or Western blotting). The transgene can be selectively expressed in some tissues of the plant or at some developmental stages, or the transgene may be expressed in substantially all plant tissues, substantially along its entire life cycle. However, any combinatorial expression mode is also applicable.

The present disclosure also encompasses seeds of the transgenic plants described above wherein the seed has the transgene or gene construct. The present disclosure further

encompasses the progeny, clones, cell lines or cells of the transgenic plants described above wherein said progeny, clone, cell line or cell has the transgene or gene construct.

An important factor in the administration of polypeptide compounds, such as an engineered I-OnuI endonuclease, or homologue thereof, and a vector encoding an engineered I-OnuI, or homologue or hybrid thereof, is ensuring that the polypeptide or vector construct has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Proteins and other compounds such as liposomes have been described and are known to the skilled artisan, which have the ability to translocate polypeptides such as an engineered I-OnuI endonuclease, or homologue or hybrid thereof, across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules (called "binary toxins") are composed of at least two parts: a translocation/binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the engineered I-OnuI endonuclease, or a homologue or hybrid thereof, and the translocation sequence. Any suitable linker can be used, *e.g.*, a peptide linker.

The variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, and constructs encoding the variant or engineered I-OnuI endonuclease or homologue or hybrid thereof can also be introduced into an animal cell, preferably a mammalian cell, via a liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate

an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, *i.e.*, a variant or engineered I-OnuI endonuclease or homologue thereof or vector encoding the I-OnuI endonuclease or homologue or hybrid thereof. The liposome fuses with the plasma membrane, thereby releasing the variant or engineered I-OnuI endonuclease or homologue or hybrid thereof into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (in this case, the engineered I-OnuI endonuclease, or homologue or hybrid thereof) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active drug release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane. When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents.

Such liposomes typically comprise a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, and a lipid component, *e.g.*, a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (*e.g.*, an antigen). A variety of methods are available for preparing liposomes, and are well known in the art. Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are known to those of skill in the art.

In certain embodiments, it is desirable to target liposomes using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (*e.g.*, ligands, receptors, and monoclonal antibodies) has been and methods for their construction and administration are well known to the skilled artisan.

5 Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes of lipid components, *e.g.*, phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A.

10 The dose of a variant or engineered I-OnuI endonuclease, or a homologue or hybrid thereof, administered to a patient, or to a cell which will be introduced into a patient, in the context of the present disclosure, should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, *e.g.*, in functional genomics
15 studies, and in cell or animal models. The dose will be determined by the efficacy and K_d of the particular variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany
20 the administration of a particular compound or vector in a particular patient.

The maximum therapeutically effective dosage of a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, for approximately 99% binding to target sites is calculated to be in the range of less than about 1.5×10^5 to 1.5×10^6 copies of the specific variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof molecule per
25 cell. The appropriate dose of an expression vector encoding a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, can also be calculated by taking into account

the average rate of expression or the variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, from the promoter and the average rate of degradation of the variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, in the cell. In certain embodiments, a weak promoter such as a wild-type or mutant HSV TK promoter is used, as described above. The dose of variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, in micrograms is calculated by taking into account the molecular weight of the particular engineered I-OnuI endonuclease, or homologue or hybrid thereof, being employed.

In determining the effective amount of the I-OnuI endonuclease, variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, to be administered in the treatment or prophylaxis of disease, the physician evaluates circulating plasma levels of the I-OnuI endonuclease, or homologue or hybrid thereof, or nucleic acid encoding the variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, potential variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, toxicities, progression of the disease, and the production of anti- I-OnuI endonuclease, or homologue or hybrid thereof, antibodies. Administration can be accomplished via single or divided doses.

Pharmaceutical compositions and administration of a variant or engineered I-OnuI endonuclease, or homologue hybrid thereof, and expression vectors encoding a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, can be administered directly to the patient for targeted single strand cleavage and/or recombination, and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by I-OnuI endonuclease, or variant or homologue or hybrid thereof, gene modification include pathogenic bacteria, *e.g.*, *Chlamydia*, rickettsial bacteria, mycobacteria, staphylococci,

streptococci, pneumococci, meningococci and gonococci, *Klebsiella*, *Proteus*, *Serratia*, *Pseudomonas*, *Legionella*, *Diphtheria*, *Salmonella*, bacilli, *Cholera*, tetanus, botulism, anthrax, plague, leptospirosis, Lyme disease bacteria and the like; infectious fungus, *e.g.*, *Aspergillus*, *Candida* species and the like; protozoa such as sporozoa (*e.g.*, *Plasmodia*), rhizopods (*e.g.*, *Entamoeba*) flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, and the like), and the like; viral diseases, *e.g.*, hepatitis (A, B, or C), herpes virus (*e.g.*, VZV, HSV-1, HSV-6, HSV-11, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, coronavirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, and the like.

Administration of therapeutically effective amounts is by any of the routes normally used for introducing a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, or an expression vector encoding a variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, of the invention into ultimate contact with the tissue or cell type to be treated. The variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, is administered in any suitable manner, preferably with a pharmaceutically acceptable carrier. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions that are available (see, *e.g.*, Remington The Science and Practice of Pharmacy, 21st ed., 2005, Lippincott Williams & Wilkins).

The variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, alone or in combination with other suitable components, can be made into an aerosol formulation (*i.e.*, "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The disclosed compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally.

The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

The disclosed methods and variant or engineered I-OnuI endonuclease, or homologue or hybrid thereof, compositions for targeted cleaving one strand of a polynucleotide sequence can be used to induce mutations in a genomic sequence, *e.g.*, by cleaving the DNA in the region of its genomic target sequence and initiating enzymatic events and subsequent mechanisms in the cell that lead to gene conversion and repair shifted to conservative, templated recombination pathways. The same methods can also be used to replace a wild-type sequence with a mutant sequence, or to convert one allele to a different allele.

Targeted DNA cleavage of an infecting or integrated viral genome can be used to treat viral infections in a host. Additionally, targeted DNA cleavage of a gene encoding a receptor for a virus can be used to block expression of such receptors, thereby preventing

viral infection and/or viral spread in a host organism. Targeted mutagenesis of a gene encoding a viral receptor can be used to render the receptor unable to bind to virus, thereby preventing new infection and blocking the spread of existing infections. Non-limiting examples of viruses or viral receptors that may be targeted include herpes simplex virus (HSV), such as HSV-1 and HSV-2, varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV), HHV6 and HHV7. The hepatitis family of viruses includes hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV). Other viruses or their receptors may be targeted, including, but not limited to, Picornaviridae (*e.g.*, polioviruses, and the like); Caliciviridae; Togaviridae (*e.g.*, rubella virus, dengue virus, and the like); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (*e.g.*, rabies virus, and the like); Filoviridae; Paramyxoviridae (*e.g.*, mumps virus, measles virus, respiratory syncytial virus, and the like); Orthomyxoviridae (*e.g.*, influenza virus types A, B and C, and the like); Bunyaviridae; Arenaviridae; Retroviridae; lentiviruses (*e.g.*, HTLV-I; HTLV-II; HIV-1, HIV-II); simian immunodeficiency virus (SIV), human papillomavirus (HPV), influenza virus and the tick-borne encephalitis viruses. See, *e.g.*, Fundamental Virology, 2nd Edition (Knipe *et al.*, eds. 2001), for a description of these and other viruses.

In similar fashion, the genome of an infecting bacterium can be mutagenized by targeted DNA cleavage followed by templated recombination, to block or ameliorate bacterial infections. The disclosed methods for targeted homologous recombination can be used to replace any genomic sequence with a homologous, non-identical sequence. For example, a mutant genomic sequence can be replaced by its wild-type counterpart, thereby providing a method for treatment of, *e.g.*, a genetic disease, an inherited disorders, cancer, and an autoimmune disease. In like fashion, one allele of a gene can be modified using the methods of targeted recombination disclosed herein.

Exemplary genetic diseases include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No. 102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangiectasia, Barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, Canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercun's disease, ectodermal dysplasia, Fanconi anemia, fibrodysplasia ossificans progressiva, fragile X syndrome, galactosemia, Gaucher's disease, generalized gangliosidosis (*e.g.*, GMI), hemochromatosis, the hemoglobin C mutation in the 6 codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefelter syndrome, Krabbe Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920), leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipidus, neurofibromatosis, Niemann-Pick disease, osteogenesis imperfecta, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, and X-linked lymphoproliferative syndrome (XLP, OMIM No. 308240).

Additional exemplary diseases that can be treated by targeted single DNA strand cleavage and/or targeted templated homologous recombination of the invention include acquired immunodeficiencies, lysosomal storage diseases (*e.g.*, Fabry disease), mucopolysaccharidosis (*e.g.*, Hunter's disease), hemoglobinopathies and hemophilias. In

certain cases, alteration of a genomic sequence in a pluripotent cell (*e.g.*, a hematopoietic stem cell) is desired. Methods for mobilization, enrichment and culture of hematopoietic stem cells are known in the art. Treated stem cells can be returned to a patient for treatment of various diseases including, but not limited to, SCID and sickle-cell anemia.

5 In many of these cases, a region of interest comprises a mutation, and the donor polynucleotide comprises the corresponding wild-type sequence. Similarly, a wild-type genomic sequence can be replaced by a mutant sequence, if such is desirable. For example, overexpression of an oncogene can be reversed either by mutating the gene or its control sequences with sequences that support a lower, non-pathologic level of expression. Any
10 pathology dependent upon a particular genomic sequence, in any fashion, can be corrected or alleviated using the methods and compositions disclosed herein.

 Targeted DNA cleavage and targeted template recombination can also be used to alter non-coding sequences (*e.g.*, regulatory sequences such as promoters, enhancers, initiators, terminators, splice sites) to alter the levels of expression of a gene product. Such methods
15 can be used, for example, for therapeutic purposes, functional genomics and/or target validation studies.

 The variant or engineered I-OnuI, and homologues and hybrids thereof, compositions and methods described herein also allow for novel approaches and systems to address immune reactions of a host to, for example, allogeneic grafts. In particular, a major problem
20 faced when allogeneic stem cells (or any type of allogeneic cell) are grafted into a host recipient is the high risk of rejection by the host's immune system, primarily mediated through recognition of the Major Histocompatibility Complex (MHC) on the surface of the engrafted cells. The MHC comprises the HLA class I protein (s) that function as heterodimers that are comprised of 3 common subunits and a variable subunit. It has been
25 demonstrated that tissue grafts derived from stem cells that are devoid of HLA escape the host's immune response. Using the variant or engineered I-OnuI, and homologues and

hybrids thereof, compositions and methods described herein, genes encoding HLA proteins involved in graft rejection can be cleaved, mutagenized or altered by templated recombination, in either their coding or regulatory sequences, so that their expression is blocked or they express a non-functional product. For example, by inactivating the gene
5 encoding the common β subunit gene (β 2 microglobulin) using a variant or engineered I-OnuI, and homologues thereof, as described herein, HLA class I can be removed from the cells to rapidly and reliably generate HLA class I null stem cells from any donor, thereby reducing the need for closely matched donor/recipient MHC haplotypes during stem cell grafting.

10 Inactivation of a gene (*e.g.*, the β 2 microglobulin or other gene) can be achieved, for example, by a single cleavage event, by cleavage followed by templated recombination, by targeted recombination of a missense or nonsense codon into the coding region, or by targeted recombination of an irrelevant sequence (*i.e.*, a "stuffer" sequence) into the gene or its regulatory region, so as to disrupt the gene or regulatory region.

15 Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art
20 from this detailed description.

Examples

The crystal structure and biochemical characterization of I-OnuI.

An endonuclease encoded within a group I intron in the RPS3 host gene from *Ophiostoma novo-ulmi* was identified as potentially displaying the characteristics of an
25 LAGLIDADG homing endonuclease required for gene targeting and subsequent engineerability. The protein (SEQ ID NO: 35) was a monomer and was found in a

mesophilic fungal host. The open reading frame of I-OnuI was inserted between BamHI and NotI sites of a pGEX-6p 3expression vector (GE Healthcare Life Sciences), and the GST fusion recombinant proteins were expressed in *E. coli* strain BL21-CodonPlus (DE3)-RIL (Agilent Technologies). Protein expression was induced in LB medium supplemented with 0.2 % glucose, 1 mM MgSO₄, and 100 µg/ml ampicillin at 16°C for about 20 hours after the culture had achieved early log growth phase (OD₆₀₀ = ~0.6). The harvested cells were resuspended in TDG buffer (20 mM Tris-HCl (pH 7.5), 1 mM dithiothreitol (DTT), and 5% glycerol) supplemented with 0.5 M NaCl. After adding lysozyme to 0.5 mg/ml, the cells were sonicated for 30 sec 6 times, and stirred on ice for 30 minutes. The clarified cell lysate was obtained by centrifugation at 25,000 x g for 30 minutes at 4 °C, and nucleic acids were precipitated by adding polyethylenimine (pH 7.9) to 0.25 % (v/v). After centrifugation at 25,000 x g for 10 minutes at 4°C, the supernatant was filtered through a 0.45 µm PVDF membrane, and mixed with glutathione sepharose 4B beads (GE Healthcare Life Sciences). The beads were extensively washed with TDG buffer supplemented with 2 M NaCl, and equilibrated with Digestion buffer (50 mM Tris-HCl (pH 7.0), 0.5 M NaCl, 1 mM dithiothreitol (DTT), and 5% glycerol. The intact I-OnuI and subsequent variant proteins were eluted by incubation with a GST tag specific protease (PreScission® protease (GE Healthcare Life Sciences) for 16 hours at 4 °C. The collected proteins were concentrated and stored at -80 °C until use as a GST fusion protein, and subsequently purified by affinity chromatography (FIGURE 3). The protein was assayed for binding and cleavage of its putative target site (corresponding to the intron insertion site in the RP3 gene of its biological genomic host) and found to display robust cleavage activity with an approximate dissociation constant (K_d) of 3 pM. To assay for the binding of DNA target site by I-OnuI each reaction mixture contained 20 mM Tris-acetate (pH 7.5), 40 mM NaCl, 1 mM CaCl₂, 1 mM DTT, 0.2 mg/ml BSA, 5% glycerol, I-OnuI , unlabeled T7 terminator primer and 10 pM radiolabeled DNA substrate containing the I-OnuI target sequence. After the reaction mixtures were

incubated on ice for 10 minutes, the protein-DNA complexes were separated on a 5% polyacrylamide gel containing 20 mM Tris-acetate (pH 7.5) and 1 mM CaCl_2 . Gel images were taken with an imager (Typhoon Trio Multi-mode imager (GE Healthcare Life Sciences)). The DNA bands were quantified using the software application ImageJ®, and the dissociation constant (KD) values were calculated using GraphPad® Prism 5 software.

Cleavage of the DNA target site by I-OnuI was determined as follows. Each reaction mixture contained 20 mM Tris-acetate (pH 7.5), 140 mM potassium glutamate (pH 7.5), 10 mM NaCl, 1 mM MgCl_2 , 1 mM DTT, 0.2 mg/ml BSA, I-OnuI or the variants, and 10 pM of the radiolabeled substrate used for the electrophoresis mobility shift assay. The reactions ran at 37°C for 30 min, and were terminated by adding 4x Stop solution (40 mM Tris-HCl (pH 7.5), 40 mM EDTA, 0.4% SDS, 10% glycerol, 0.1% bromophenol blue and 0.4 mg/ml proteinase K). After incubation at 37°C for 15 minutes, each sample was loaded on a 20% polyacrylamide-TBE gel. The gel images were taken and the DNA bands were quantified as described above.

The purified enzyme was crystallized in complex with its target site, and the structure of the protein-DNA complex was determined to 2.4 Å resolution (FIGURE 4). The structure demonstrated that the enzyme forms a monomeric LAGLIDADG (LHE) protein fold with two pseudo-symmetric DNA-binding domains (LD1 amino acid residues 13 to 24 of SEQ ID NO: 35 and amino acid residues 54 to 66 of the consensus sequence shown in Figure 11; LD2 amino acid residues 160 to 171 of SEQ ID NO: 35 and amino acid residues 238 to 249 in Figure 11), connected by a flexible linker (amino acid residues 141 to 156 of SEQ ID NO: 35 and amino acid residues 202 to 234 of Figure 11). The LHE domains form an elongated protein fold that consists of a core fold with mixed α/β topology (α - β - β - α - β - α). The overall shape of this domain is a half-cylindrical "saddle" that averages approximately 25 Å x 25 Å x 35 Å. The surface of the saddle is formed by an anti-parallel β -sheet within each protein domain that presents a large number of exposed basic and polar residues for

DNA contacts and binding. The complete DNA-binding surfaces of the full-length enzymes, generated by two-fold pseudo-symmetry, was approximately 80 Å long and accommodated the DNA target site.

The LAGLIDADG motifs of I-OnuI form the last two turns of the N-terminal helices in each folded domain or monomer that are packed against one another. They also contribute N-terminal, conserved acidic residues (E22 and E178) to two active sites where they coordinate divalent cations that are essential for catalytic activity. The structure and packing of the parallel, two-helix bundle in the domain interface of the LAGLIDADG enzymes are strongly conserved among the diverged members of this enzyme family.

10 **The recognition of a DNA target sequence by I-OnuI.**

Amino acid side chains from the two structurally independent, antiparallel β-sheets (one from each protein domain) are used to contact nucleotide bases within the major groove, at positions flanking the central four base pairs as shown in FIGURE 5. The 22 basepair target site, as shown in the illustration, is recognized by a combination of at least 22 direct contacts between amino acid side chains and individual DNA bases, at least 14 additional water-mediated contacts between amino acid side chains and individual DNA bases, and at least 30 contacts between the protein and the DNA phosphoribose backbone (mostly water-mediated). At least 40 amino acid residues are involved in direct or water-mediated contacts to the DNA target; these represent the first shell of amino acid residues that can be exploited for the redesign or selection, as described below. The roles of these amino acids are not mutually exclusive: the same side chain can be used for different forms of readout of adjacent DNA bases or backbone atoms.

The DNA binding and cleavage specificity profile of I-OnuI.

The DNA binding and DNA cleavage specificity profile of surface-displayed I-OnuI was determined as described in Jarjour *et al.* (*Nucl. Acids Res.* 37:6871-6880, 2009, incorporated herein by reference in its entirety) (FIGURE 6). The cleavage specificity

profile of I-OnuI using a yeast surface display cleavage assay. The cleavage specificity profile of I-OnuI was obtained by measuring the relative cleavage activities of 66 target sequences, each containing a single base-pair substitution from the enzyme's original target site. Cleavage activity against each DNA target was measured using yeast surface-displayed enzyme and flow-cytometry-based tethered DNA cleavage assays. Approximately 100,000 cells expressing I-OnuI were stained with 1:250 dilution biotinylated antibody against hemagglutinin (HA)-epitope tag (Covance) and 1:100 fluorescein isothiocyanate (FITC)-conjugated α Myc (ICL Labs) for 30 minutes at 4°C in 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose. The cells were then stained with pre-conjugated streptavidin-PE:Biotin-dsOligo-A467 in the same buffer supplemented with 400mM KCl. The cells were washed in the buffer containing 180 mM KCl, and split into two wells. Each well was then resuspended in the same buffer supplemented with 2mM of either $MgCl_2$ or $CaCl_2$. After incubation at 37°C, the cells were pelleted and resuspended in the buffer containing 400 mM KCl and 4mM EDTA to enhance release of the cleaved substrates, and analyzed on a BD LSRII cytometer. The relative cleavage activities for each DNA target site were evaluated by calculating the ratio of the non-cleavage median DNA-Alexa647 fluorescence intensity to the post-cleavage intensity in the matching gate.

The binding specificity profile of I-OnuI was determined using a yeast surface display binding assay. The binding specificity profile of I-OnuI was obtained by measuring the relative binding of 66 target sequences, each containing a single base-pair substitution from the enzyme's original target site. The relative binding of each DNA target was measured using yeast surface-displayed enzyme in and flow-cytometry-based untethered DNA retention binding assays. Approximately 100,000 cells expressing I-OnuI were stained with a serial titration of pre-conjugated streptavidin-PE:Biotin-dsOligo-A467 in 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose and 2 mM $CaCl_2$. After incubation at 37°C, the cells were washed in the same buffer containing 180

mM KCl and the retained, bound DNA was measured by direct fluorescence of the A467 label on the DNA.

The specificity profiles illustrate the ability of the enzyme to bind and cleave a series of alternative DNA target sequences that each contain a single basepair mismatch at each of the 22 positions in the DNA recognition site. The analysis indicates that the majority of specificity of DNA recognition is accomplished during DNA binding, rather than at the chemical step of DNA hydrolysis (although there are several individual basepair substitutions that do not affect binding, but inhibit subsequent cleavage). Those basepair substitutions that can be bound and cleaved as well (or at least with an efficiency greater than 50% that of the corresponding wild-type DNA base) include -6C, -4C, +4T and +4C. The overall specificity of the enzyme is therefore extremely high (at least 1 in 10^{10}).

Design and selection of I-OnuI, I-OnuI homologues and hybrids thereof with altered DNA binding specificity at individual basepairs of its naturally occurring target site.

Variants of I-OnuI, including variants of I-OnuI and homologues and hybrids thereof, that can cleave altered DNA target sites (containing individual basepair substitutions) were identified using a combination of structure-based design and genetic selection for cleavage activity. The method relied upon identification of amino acid contact residues near each altered DNA basepair that were identified in the protein-DNA crystal structure (illustrated in FIGURE 5) and subsequent selection of surrounding amino acid substitutions (corresponding to a 'pocket' of protein side chains that surround the immediate contacting residue at each position in the protein-DNA interface). The analysis identified alternative amino acid identities that can form contacts to altered basepairs at each position in the DNA target, while modeling a conservatively flexible nucleotide and protein backbone. Subsequent creation of limited protein mutation libraries and screening these libraries for DNA cleavage that leads to elimination of a bacterial reporter gene has produced active enzyme variants with desired

specificities for many single base-pair substitutions in the wild-type target site of I-OnuI homing endonuclease (FIGURE 7, bold amino acid residues). Specifically, desired I-OnuI sequences were mapped on to the I-OnuI-DNA crystal structure and the Rosetta computational design methodology was used to optimize the amino acid sequence of the protein to maximize affinity for the new site. The predicted specificity of the resulting protein models for the desired target sequence was computed using Rosetta, and designs that were predicted to bind tightly and specifically were subjected to further optimization using flexible backbone protein design. The tightest binding and most specific designs were again selected, and the designed amino acid substitutions were removed one at a time. If no significant loss was predicted in either specificity or binding energy, the substitution was removed from the design. Genes encoding the designed proteins were assembled from oligonucleotides, and the designed proteins were expressed, purified, and assayed as described above.

The selection system was also modified to allow for selection of enzyme specificity as well as activity by selecting against enzyme variants that can still cleave the wild-type target site. In one particular embodiment, the open reading frames of I-OnuI was inserted between NcoI and NotI sites of the pEndo expression vector. Expression of the I-OnuI gene was tightly regulated by the pBAD promoter, and addition of L-arabinose promoted gene transcription. Site-directed mutagenesis or random mutagenesis on I-OnuI gene was induced both by overlap extension PCR, and by using GeneMorph II Random Mutagenesis Kit (Agilent Technologies), by following the manufacture's instructions. Two copies each of a desired I-OnuI target site were inserted between AflIII and BglII and between NheI and SacII of the reporter "pCcdB" plasmid. This plasmid encodes "control of cell death B" protein, which is toxic to bacteria; the expression of this gene is induced by isopropyl- β -thiogalactopyranoside (IPTG). The cleavage of the target sites by the expressed LHEs leads to the RecBCD-mediated degradation of the reporter plasmid, and rescues the cell growth on

the selective medium containing IPTG. The pEndo plasmid was transformed into NovaXGF' (Novagen) competent cells harboring pCcdB plasmid (containing four copies of the I-OnuI or MAO-B target) by electroporation. The transformants were grown in 2 x YT medium at 37 °C for 30 minutes, and were 10-fold diluted with 2 x YT medium supplemented with 100 mg/ml carbenicillin and 0.02% L-arabinose. After the culture was grown at 30°C for 4-15 hours, the cells were harvested, resuspended in sterilized water, and spread on both non-selective plates (1 x M9 salt, 1% glycerol, 0.8% tryptone, 1 mM MgSO₄, 1 mM CaCl₂, 2 mg/ml thiamine, and 100 mg/ml carbenicillin) and selective plates (the non-selective plates supplemented with 0.02% L-arabinose and 0.4 mM IPTG). For negative selection to remove the variants active against the wild type target sequence, the transformants were spread on the selective plates containing 33 mg/ml chloramphenicol instead of IPTG. The plates were incubated at 30°C for about 30 to 40 hours. To proceed to the next round of selection, the pEndo plasmid was extracted from the surviving colonies on the selective plates. The ORFs of I-OnuI variants were recovered by PCR amplification, and digested with NcoI, NotI, and ScaI or PvuI. The resulting fragments were cloned into pEndo vector, and subjected to further selection.

Engineering of a variant of I-OnuI against a physiological target site in the human genome (the monoamine oxidase B (MAO-B) gene).

A variant of the I-OnuI endonuclease has been engineered that targets a unique sequence in the MAO-B target site, and that can be used to direct homologous recombination in that gene locus, leading to targeted point mutations of the endogenous MAO-B gene and development of engineered neural cell lines for *in vivo* studies.

Design and selection of the MAO-B specific variant of I-OnuI was carried out by identification and mutation of residues in the protein -DNA interface that contact individual basepairs that differ between the desired human target site (MAO-B) and the wild-type I-OnuI recognition site (FIGURE 9). The mutation and redesign process was carried out in a

sequential, iterative manner, by isolating enzyme mutations that could accommodate and cleave DNA variants with basepair substitutions corresponding to -10G, -4A and +2T, respectively. Directed evolution of I-OnuI endonuclease to cleave the human MAO-B gene target was carried out using the *in vivo* bacterial screening methodology described for Figure 5 7, and a scheme for mutagenesis in which all residues within 5 angstrom contact distance of the altered base pairs in the target site (N32, S40, T48, S72, K80, K229, W234, and D236) were randomized, and selections were carried out in sequential order of addition as shown in rounds 1, 2 and 3 of the figure. Addition of an E178D substitution into the active site of E1 I-OnuI increased cleavage activity of the enzyme and was included in the final redesigned 10 enzyme construct (E178 is one of two residues that coordinate divalent metal ions in the active site. Electrophoretic mobility shift assays using purified recombinant proteins, as described above, demonstrated that wild-type I-OnuI preferentially bound its physiological target with a very tight dissociation constant (193 ± 15.2 pM). The E1 and E2 I-OnuI proteins displayed similar affinity for both the WT and MAO-B targets; however these 15 enzymes significantly discriminated between the two target sites in cleavage reactions. The relative cleavage activities assayed *in vitro* correlated well with the GFP gene conversion frequencies that were measured using the DR-GFP reporter. For instance, E2 I-OnuI induced GFP gene conversion on the MAO-B target approximately 3-fold more efficiently than E1 I-OnuI, and displayed a very similar level of *in vitro* cleavage activity for the MAO-B target at 20 approximately 4 fold lower enzyme concentrations. The final variants of I-OnuI that displayed altered cleavage specificity towards the desired human MAO-B target site (denoted R3 #3, R3#6 and R3#8 in FIGURE 9) each harbored six amino acid substitutions: N32L, S40R, T48M, S27R, K80R, and K229R, H, or Y). The relative binding and cleavage activity of wild-type and "R3 #3" variant of I-OnuI towards the wild-type and human 25 MOA-B genomic target site were assayed (FIGURE 10). The relative binding by wild-type (WT) and redesigned (R3#3) I-OnuI towards the original (WT) and MAO-B DNA target

sites were performed using electrophoretic mobility shift assays and the relative cleavage of by wild-type (WT) and redesigned (R3#3) I-OnuI towards the original (WT) and MAO-B DNA target sites were performed using *in vitro* enzyme cleavage digests as described above. In addition, the relative cleavage of by wild-type (WT) and redesigned (R3#3) I-OnuI towards the original (WT) and MAO-B DNA target sites were performed using *in vitro* enzyme cleavage digests as described above. Sequences of human MAO-B gene targets before and after treatment with redesigned I-OnuI enzyme were determined by extracting genomic DNA from sorted cells ($\sim 1 \times 10^5$) which had been washed with cold PBS buffer, resuspended in TNES buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM EDTA, 1% SDS, 0.25 mg/ml proteinase K), and incubated at 50°C for 30 minutes. RNase A was added to 0.25 mg/ml, and the reaction mixture was further incubated at the same temperature for 30 minutes. The genomic DNA was recovered by phenol/chloroform/isoamyl alcohol (PCI) extraction followed by ethanol precipitation. Both of the on-target (*i.e.*, MAO-B gene) and off-target loci were amplified from 50-80 ng of the extracted genomes using Phusion DNA polymerase (Finnzymes). The DNA products resulting from 2 rounds of PCR amplification were cleaned using a PCR purification kit (Qiagen). Individual clones were sequenced using dye-terminator Sanger sequencing on an ABI automated DNA sequencer

In addition, small deletions were found within the endogenous MAO-B target site during targeted mutagenesis of the endogenous MAO-B gene in human tissue culture cells. Table 1. The intact genome sequence of the MAO-B gene, prior to treatment with the R3 #3 variant of I-OnuI is shown at the top of the Table and the MAO-B target sequence is underlined; the mutated sequences of the same gene caused by treatment with the R3 #3 variant of I-OnuI are aligned below.

Table 1. Deletions found in the endogenous MAO-B gene using the R3#3 variant I-OnuI for targeted mutagenesis.

SEQ ID NO:110	CTGGGTTGGTCCAACATAGGATCCTCCAA <u>GGTCCACATATTTAACCTTTTGGT</u> CTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:111	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCACAT-TTTAACCTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:112	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCACATATT--CCTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:113	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCACA-----AACCTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:114	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCACATA-----TTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:115	CTGGGTTGGTCCAACATAGGATCCTCCAA-----ATTAACTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:116	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCA-----CCTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA
SEQ ID NO:117	CTGGGTTGGTCCAACATAGGATCCTCCAAGGTCCACATATTTA-----TTTCCCATAGGAAAAAATTAAA
SEQ ID NO:118	CTGGGTTGGTCCAACATAGGATCCTCCAA-----CCTTTTGGTCTGTTTCCCATAGGAAAAAATTAAA

Methods for knockout of the MAO-B gene

A plasmid construct to express engineered I-OnuI (pExodus®) was created in which the gene including the N-terminal hemagglutinin (HA) tag, followed by a nuclear localization signal was linked to an mCherry gene by the 2A peptide sequence from *Thosea* 5 *asigna* virus (T2A). The two-gene expression was driven by a cytomegalovirus (CMV) promoter, and the co-translated proteins were separated by ribosomal skipping. The DR-GFP reporter codes a GFP gene sequence interrupted by a HE target site and an in-frame stop codon, followed by the truncated gene sequence. Human embryonic kidney (HEK) 293T cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 10 units/ml penicillin and 10 mg/ml streptomycin at 37°C in 5% CO₂ 10 atmosphere. 6 x 10⁴ of HEK 293T cells were plated 24 h prior to transfection in 12-well plates, and transfected with 0.25 mg each of DR-GFP reporter and pExodus® plasmid using a transfection reagent (Fugene® 6 transfection reagent (Roche Applied Science)). The GFP positive cells were detected by flow cytometry at 48 h post transfection. Western blotting 15 was carried out using rabbit polyclonal antibody against hemagglutinin (HA)-epitope tag and mouse monoclonal antibody against β-actin.

HEK 293T cells (1.3 x 10⁵) were plated 24 hours prior to transfection in 6-well plates, and transfected with 1 mg of pExodus® plasmid. The top 25% and the following 25% of mCherry positive cells (fluorescent marker for a LHE gene expression) were separately 20 collected using BD FACSaria® cell sorter (BD Biosciences) 48 hours post transfection. To extract genomic DNA, the sorted cells (~ 1 x 10⁵) were washed with cold PBS buffer, resuspended in TNES buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM EDTA, 1% SDS, 0.25 mg/ml proteinase K), and incubated at 50°C for 30 minutes. RNase A was added to 0.25 mg/ml, and the reaction mixture was further incubated at the same temperature for 30 25 minutes. The genomic DNA was recovered by phenol/chloroform/isoamyl alcohol (PCI) extraction followed by ethanol precipitation. Both of the on-target (*i.e.*, the MAO-B gene)

and off-target loci were amplified from 50-80 ng of the extracted genomes using a DNA polymerase (Phusion® DNA polymerase, Finnzymes). The DNA products resulting from 2 rounds of PCR amplification were cleaned using a PCR purification kit (Qiagen), and 150 mg of the fragments were incubated with 1.5-3.0 pmol of E2 I-OnuI recombinant protein in 5 20 mM Tris-acetate (pH 7.5), 100 mM potassium acetate (pH 7.5), 1 mM DTT and 10 mM MgCl₂ at 37 °C for 2 hours. The cleavage reactions were terminated by adding 4 x Stop solution (40 mM Tris-HCl (pH7.5), 40 mM EDTA, 0.4% SDS, 10% glycerol, 0.1% bromophenol blue and 0.4 mg/ml proteinase K). After incubation at 37°C for 30 minutes, each sample was separated on a 1.8% agarose gel containing ethidium bromide in TBE. The 10 DNA bands were quantified using ImageJ® software. The MAO-B gene was successfully knocked out using this method.

The identification and modeling of structural and functional homologues of I-OnuI.

Recent high-throughput sequencing of Ascomycete fungi have resulted in the discovery of unique lineages of LAGLIDADG homing endonucleases inserted in a variety of 15 host genes. (Sethuraman *et al.*, *Mol. Biol. Evol.* 26:2299-2315, 2009). These endonucleases display a wide repertoire of new specificities. At least 26 recognizable homologues of I-OnuI, displaying an average of 45% sequence identity, similar overall peptide chain lengths, and at least 50% sequence identity of amino acid sequence within the two LAGLIDADG domains and the "Loop sequence" as described above, relative to that 20 enzyme, have been identified and aligned (FIGURE 11). Using this data and the crystal structure of I-OnuI, structure-based homology models of these homing endonucleases have been generated with a corresponding list of predicted DNA-contacting residues for each enzyme.

Methods for selection of homologues and alignment to I-OnuI

25 Putative full-length LAGLIDADG homing endonuclease that displayed homology to the I-OnuI homing endonuclease were collected from the Pfam database, available on the

world-wide-web and described in Finn *et al.*, *Nucl. Acid Res.* 38:D211-D222, 2010, based on an observed value at least 40% amino acid sequence identity to the I-OnuI protein. Structure-based multiple sequence alignments were then built using Cn3D application available at the National Center for Biotechnology Information website. After the structure of I-OnuI was aligned to the structure of I-AniI (PDB 1P8K) the collected Pfam sequences were aligned to the I-OnuI and I-AniI structures, at which point homodimeric LAGLIDADG sequences were removed based on length and the occurrence of a single LAGLIDADG motif per sequence. The sequence alignment generated by the Cn3D application was subsequently validated using a modified version of the Java based multiple alignment editor application Jalview that calculates the MIp/Zp co-variation statistic in real-time while the alignment is edited. Groups of misaligned sequences were realigned to minimize local co-variation, as local co-variation is a unique indicator of misalignment that is independent of methods used to build the multiple sequence alignment. Local co-variation was also used as a guide to reject partial and erroneous sequences.

Methods for solving the crystal structure of I-LtrI

To express and purify I-LtrI, 10 ml of an *E. coli* culture containing a pET plasmid with the I-LtrI endonuclease (pET-15HE-Ltr) was grown overnight and diluted 1:100 into 1 liter of Luria-Bertani media. The 1 liter culture was grown at 37°C for 3 hours, shifted to 27°C, and expression induced by adding isopropyl-b-D-thiogalactopyranoside to a final concentration of 1 mM. After additional growth for 2.5 hours, cells were harvested by centrifugation at 5000 rpm for 5 min and the pellet was frozen at -80 °C. For protein purification, the frozen cells were thawed in the presence of protease inhibitor (Roche Diagnostic) and resuspended in 10 ml of lysis buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 40 mM imidazole and 10% glycerol) per 1 gm of wet cell weight. Cells were disrupted by homogenization followed by centrifugation at 27,200 g for 25 min at 4 °C. The supernatant was sonicated to facilitate DNA fragmentation, and centrifuged at 20,400 g for

15 min at 4°C. The supernatant was applied to a HisTrap HP® Affinity column (GE Healthcare) that had been charged with 0.1 M NiSO₄ and equilibrated with binding buffer (20 mM Tris-HCl, pH7.9, 500 mM NaCl, 40 mM imidazole, and 10% glycerol). Bound protein was eluted with elution buffer (20 mM Tris-HCl, pH7.9, 500 mM NaCl, and 10% glycerol) over a linear gradient of imidazole from 0.08 to 0.5 M, and 0.5 mL fractions were collected over 50 ml. To prevent precipitation, 500 microliters of 2 M NaCl and 10 microliters of 0.5 M EDTA, pH8.0, were added to peak fractions. The peak fraction was loaded directly onto a Superdex 75 gel-filtration column (GE Healthcare) equilibrated with lysis buffer without imidazole. Fractions were collected in 0.25-ml aliquots over 25 ml.

To obtain I-LtrI-DNA co-crystals, the DNA oligonucleotides (5'-GGTCTAAACGTCGTATAGGAGCATTTGG-3' (SEQ ID NO: 119) and 5'-CAAATGCTCCTATACGACGTTTAGACCC-3' (SEQ ID NO:120)) were purchased from Integrated DNA Technologies (1 mmole scale, standard desalting purification). The oligonucleotides were dissolved in TE buffer (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA), and the complementary DNA strands were annealed by incubation at 95°C for 10 min and slow cooling to 4°C over a six hour period. One hundred mM I-LtrI protein in 50 mM Hepes-NaOH (pH 7.5), 150 mM NaCl, 5mM MnCl₂ and 5 % (v/v) glycerol was mixed with 1.5-fold molar excess of the DNA substrate. The protein-DNA drops were mixed in a 1:1 volume ratio with a reservoir solution containing 100 mM Bis-Tris (pH 6.5), 200 mM magnesium chloride, and 20 % (v/v) polyethylene glycol 3500 and equilibrated at 22°C. The crystals diffracted up to approximately 2.7 Å resolution at the ALS beamline 5.0.1. The data set was processed using HKL2000 package. The polyalanine model of I-OnuI/DNA complex (PDB ID: 3QQY) was used as a search model for molecular replacement. One copy of the search model was found and the structure was refined using REFMAC5. The final model was deposited in RCSB Protein Data Bank with ID code 3R7P.

Expression of I-OnuI homologues; prediction and confirmation of their individual DNA recognition sites.

The reading frames corresponding to twelve homologues of I-OnuI have been cloned and used to test expression and cleavage activity on the surface of yeast (FIGURE 13). As well, several of the same genes have been placed into inducible bacterial expression system and tested. Of twelve individual clones that have thus far been examined (CpaI, GpiI, GzeI, HjeII, LtrI, MpeI, PanI, PanII, PanIII, PnoII, ScuI, and SscI), six have displayed surface expression on yeast using the methods described in WO 2007/123636, incorporated herein by reference, and production of soluble protein in bacteria (indicating proper folding and formation of stable protein).

Target sites for each LHE were predicted through comparison of the LHE-harboring host gene to related genes lacking an endonuclease. Cleavage activity against each predicted target was verified using yeast surface-displayed enzyme in both *in vitro* and flow-cytometry-based tethered DNA cleavage assays (7) with the following modifications. Briefly, approximately 5×10^5 cells expressing a LHE were stained with 1:250 dilution biotinylated antibody against hemagglutinin (HA)-epitope tag (Covance) and 1:100 fluorescein isothiocyanate (FITC)-conjugated α Myc (ICL Labs) for 30 minutes at 4°C in 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose. The cells were then stained with pre-conjugated streptavidin-PE:Biotin-dsOligo-A467 in the same buffer supplemented with 400 mM KCl. The cells were washed in the buffer containing 180 mM KCl, and split into two wells. Each well was then resuspended in the same buffer supplemented with 2 mM of either $MgCl_2$ or $CaCl_2$. After incubation at 37°C, the cells were pelleted and resuspended in the buffer containing 400 mM KCl and 4 mM EDTA to enhance release of the cleaved substrates, and analyzed on a BD LSRII cytometer.

To identify the exact cleavage positions and overhangs that were generated for each target site, the verified sequences were cloned into the pCTCON2-ARL vector (pCTCON2

with modified cloning sites) between the NdeI and XhoI cloning sites. Intact plasmid containing the target sequence was first digested with the HindIII restriction enzyme to create a 6300 bp linear substrate. Cleavage of the target sequence in this linear substrate creates two bands (1.3 kb and 5 kb), which were clearly distinct from the uncut substrate (6.3 kb). Each *in vitro* cleavage reaction contained five million yeast expressing a LHE on the surface, 10mM DTT (to release the enzyme from the yeast surface), 5-10 micrograms of HindIII-linearized target plasmid, 5 mM MgCl₂, 10 mM Hepes (pH 7.5), 180 mM KCl, 10 mM NaCl, 0.2% BSA, and 0.1% galactose. After incubation at 37°C for 2-4 hours, the yeast cells were then spun down and the supernatant was loaded onto an agarose gel. The two product bands were purified from the gel and sequenced, using a forward primer (5'-GTTCCAGACTACGCTCTGCAGG-3', SEQ ID NO: 121) for the 5kb band, and a reverse primer (5'-GTGCTGCAAGGCGATTAAGT-3', SEQ ID NO: 122) for the 1.3kb band. The sequencing reads ended abruptly at the position of each DNA strand cleaved by the enzyme.

The predicted DNA target sites for each enzyme correspond to DNA sequences encompassing the intron or endonuclease gene insertion site in the host gene and corresponding biological genomes that is the source of the homing endonuclease reading frame (FIGURE 13). The sequence identity of these target sites, relative to that of I-OnuI, range from 41% to 91%. The target site of I-OnuI, I-LtrI, I-GpiI, and I-MpeI have been verified using standard *in vitro* cleavage assays.

Methods for Homology Modeling

Homology models were created using the SWISS-MODEL automated homology modeling server available on the world-wide-web and described in Arnold *et al. Bioinformatics* 22:195-201, 2006. Amino acid sequences for each homologue were provided as input, and the structure of either I-OnuI or I-LtrI was designated as the modeling template. Homology models for I-GpiI and I-MpeI are shown in Figure xx.

Creation of hybrid LAGLIDADG homing endonucleases via sequence substitutions between I-OnuI enzyme homologues.

Homology models of three homologues of I-OnuI (I-LtrI, I-PnoII and I-MpeI) were used to identify surface residue positions which could be mutated to create uniform protein surfaces and increased overall sequence homology between these enzymes. Residues near to the protein-DNA interface were avoided, to maintain the unique DNA-binding preferences of each individual enzyme. Multiple sequence alignments helped guide the choice of amino acid for each surface position, with the ultimate goal being both elimination of hydrophobic residues on the surface, and selection of amino acid sidechains commonly found in those homologs with the best surface expression in yeast and highest solubility and activity in bacteria. These sequence exchanges between homologous enzymes resulted in overall increased DNA coding identity (from initial values of 40% to 50%, up to 70% to 80% between the enzymes tested while maintaining enzyme activity). Therefore, 'hybrid' enzymes containing sequence elements from individual members of the I-OnuI family homologues, can be generated and used as a broad set of protein scaffolds for design and selection of additional DNA cleavage specificities.

Methods for construction of hybrid (chimeric) homing endonucleases.

N- and C-terminal half-domains of two I-OnuI homologues (I-OnuI and I-LtrI) were constructed by assembly PCR using oligonucleotides designed by the DNAsworks server (Hoover and Lubkowski, *Nucl. Acids Res.* 30:e43, 2002). Each half-domain construct was flanked by 30-50 base pairs of the pETCON vector to facilitate cloning into that expression vector via homologous recombination. The design for the genes encoding these chimeras included a Ser-Gly-Thr linker between the N- and C-terminal protein domains (which can be encoded by a DNA sequence containing a unique KpnI restriction site, which is useful for subsequent recloning and fusion of new domain combinations). The PCR product for each half-domain was purified using a PCR purification kit (Qiagen), then digested with KpnI (Fermentas) for 15 min at 22°C. Digested N-terminal and C-terminal half-domains were then

combinatorially mixed and ligated together to create genes encoding the desired full-length chimeras. In the case of chimeric enzymes lacking the synthetic Ser-Gly-Thr sequence and corresponding KpnI restriction site, the entire full-length enzyme was constructed by PCR assembly.

5 The genes encoding the full-length hybrid chimeras were combined at approximately 1:10 molar ratio with the yeast-surface display vector pETCON (Genscript), digested with XhoI and NdeI (NEB) for 4 hrs at 37°C. The digested vector and assembled enzymes were transformed into *Saccharomyces cerevisiae* strain EBY100 using the lithium-acetate protocol (Gietz and Schiestl, *Nat. Prot.* 2:38-41, 2007). Transformed yeast were grown for 3 days at
10 30°C in selection media. Clones were obtained by isolating plasmids from yeast populations using the a plasmid preparation kit (Zymoprep-II® kit; Zymo Research) and electroporating these into *Escherichia coli* DH10B (Invitrogen) for sequencing. The bacterial population harboring the correct plasmid was then grown overnight, and the clonal plasmid was isolated using a DNA miniprep kit (Qiagen); this plasmid was then transformed into EBY100 yeast
15 by LiAc protocol, as above.

Subsequent to the creation of the initial chimeric endonuclease constructs as described above, the residues that comprise the domain interface in and near the LAGLIDADG motif were randomized, and active constructs were selected. In order to create protein libraries randomized at specified LAGLIDADG interface residues,
20 oligonucleotides containing NNS codons were substituted in the PCR assembly reaction and transformed into EBY100 yeast, as above. Library size was determined by serial dilution, with typical yields of approximately 10×10^6 unique transformants. Mutation distribution and frequencies were verified by sequencing of an unselected library, and no biases were noted. Yeast were propagated in selective growth media with 2% raffinose + 0.1% glucose
25 at 30°C for 12 to 20 hours, and then induced in media with 2% galactose for 2 to 3 hrs at 30°C, followed by 16 to 24 hrs at 20°C.

Active chimeras were identified and isolated using a flow-cytometric protocol as described in Jarjour *et al.*, *Nucl. Acids Res.* 37:6871-6880, 2009. Briefly, yeast surface-displayed chimeric homing endonucleases were incubated first with a 1:300 dilution of biotinylated anti-hemagglutinin (Covance) for 30 min at 4°C in a buffer containing 180 mM KCl, 10 mM NaCl, 10 mM HEPES, 0.1% galactose, 0.2% BSA, and pH 8.25. Yeast were then washed, and incubated with pre-conjugated streptavidin-phycoerythrin(PE):biotin-DNA-Alexa fluor 647, in the same buffer as above with 580 mM KCl, for 30 min at 4°C. Cells were again washed, and transferred to a buffer containing 150 mM KCl, 10 mM NaCl, 10 mM HEPES, 5 mM K-Glu, 0.05% BSA, and pH 8.25, with 7 mM CaCl₂ or MgCl₂ for control and cleavage reactions, respectively. The yeast were incubated for 5 to 30 min at 37°C to allow catalysis; the reaction was halted by centrifugation and washed with the buffer above containing 580 mM KCl. Fluorescein isothiocyanate (FITC)-conjugated anti-Myc (ICL labs) was added to the washed cells at 1:100 dilution, and allowed to incubate for at least 10 minutes prior to flow-cytometric acquisition. See Figure 19.

Using a BD FACSaria™ II cell sorter, cells were hierarchically gated for single yeast cells surface expressing full-length enzyme. Yeast cells within these gates showing decreased Alexa-fluor 647 signal (indicating catalytic activity) were sorted using maximal phase and purity masking. Sorted yeast were expanded in culture, and analyzed for increased catalytic activity. Plasmid was isolated from yeast populations and electroporated into *E. coli* (as above) for sequencing. All data was analyzed using FloJo® software (Tree Star).

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

CLAIMS

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A method for selecting a variant or engineered I-OnuI endonuclease with a
5 target site modification from the wild-type and directed to a site within a gene of interest comprising:
 - i) determining the target site for a I-OnuI endonuclease;
 - ii) searching a nucleic acid database for a gene of interest comprising a
nucleotide sequence that is at least 40% identical to the nucleotide sequence of target site of
10 the I-OnuI endonuclease
 - iii) selecting a gene of interest comprising the nucleotide sequence that is
at least 40% identical to the nucleotide sequence of the target site of a I-OnuI endonuclease
and the I-OnuI endonuclease;
 - iv) constructing a molecular model of the I-OnuI endonuclease bound to
15 the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target
site of the I-OnuI endonuclease from the gene of interest;
 - v) mutating the I-OnuI endonuclease at amino acid residues that have
been determined to be direct contact residues, backbone contact residues, or water-mediated
contact residues with the target site of the gene of interest to form a library of variant or
20 engineered I-OnuI endonuclease;
 - vi) expressing the library of variant or engineered I-OnuI endonuclease;
 - vii) screening the library of variant or engineered I-OnuI endonuclease for
binding activity to the target sequence in the selected gene and the cleavage activity for the
target sequence in the selected gene; and
 - 25 viii) selecting the variant or engineered I-OnuI endonuclease that can act
upon a nucleotide sequence containing a modification in the target site from the wild-type

and directed to a target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

2. A method for selecting an engineered I-OnuI endonuclease homologue with a target site modification from the wild-type and directed to a site within a gene of interest comprising:

i) determining the target site for a I-OnuI endonuclease homologue;
ii) searching a nucleic acid database for a gene of interest comprising a nucleotide sequence that is at least 40% identical to the nucleotide sequence of target site of the I-OnuI endonuclease homologue;

iii) selecting a gene of interest comprising the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease homologue;

iv) constructing a molecular model of the I-OnuI endonuclease homologue bound to the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the I-OnuI endonuclease homologue from the gene of interest;

v) mutating the I-OnuI endonuclease homologue at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of engineered I-OnuI endonuclease homologues;

vi) expressing the library of engineered I-OnuI endonuclease homologues;

vii) screening the library of engineered I-OnuI endonuclease homologues for binding activity to the target sequence in the selected gene and the cleavage activity for the target sequence in the selected gene; and

viii) selecting the variant or engineered I-OnuI endonuclease homologue that can act upon a nucleotide sequence containing a modification in the target site from the

wild-type and directed to a target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

3. A method for selecting an engineered hybrid of two I-OnuI or I-OnuI homologue domains with a target site modification from the wild-type and directed to a site within a gene of interest comprising:

i) determining the target site for a hybrid of two I-OnuI or I-OnuI homologue domains;

ii) searching a nucleic acid database for a gene of interest comprising a nucleotide sequence that is at least 40% identical to the nucleotide sequence of target site of the hybrid of two I-OnuI or I-OnuI homologue domains;

iii) selecting a gene of interest comprising the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the hybrid of I-OnuI or I-OnuI homologue domains;

iv) constructing a molecular model of the hybrid of two I-OnuI or I-OnuI homologue domains bound to the nucleotide sequence of the nucleotide sequence that is at least 40% identical to the nucleotide sequence of the target site of the hybrid of two I-OnuI or I-OnuI homologue domains from the gene of interest;

v) mutating the hybrid of two I-OnuI or I-OnuI homologue domains at amino acid residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of engineered hybrids of two I-OnuI or I-OnuI homologue domains;

vi) expressing the library of engineered hybrids of I-OnuI or I-OnuI homologue domains;

vii) screening the library of engineered hybrids of I-OnuI or I-OnuI homologue domains for binding activity to the target sequence in the selected gene and the cleavage activity for the target sequence in the selected gene; and

viii) selecting the engineered hybrid of two I-OnuI or I-OnuI homologue domains that can act upon the nucleotide sequence containing a modification in the target site from the wild-type and directed to a target site within the gene of interest, wherein the binding and cleavage activity is highest for the target sequence in the gene of interest.

5 4. A method for producing an engineered I-OnuI endonuclease, an engineered I-OnuI endonuclease homologue, or a hybrid of two I-OnuI or I-OnuI homologue domains with a target site modification from the wild-type and directed to a site within a gene of interest comprising:

 i) determining the nucleotide sequence of the gene of interest;

10 ii) searching a nucleic acid database comprising the target sites for I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains for a I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains comprising a nucleotide sequence that is at least 40% identical to a nucleotide sequence of within the gene of interest;

15 iii) selecting the I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains comprising the I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains with the target site nucleotide sequence that is at least 40% identical to the nucleotide sequence within the gene of interest;

20 iv) constructing a molecular model of the selected I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains bound to the nucleotide sequence of the target site that is at least 40% identical to the nucleotide sequence within the gene of interest with the nucleic acid sequence within the gene of interest;

25 v) mutating the selected I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains at amino acid

residues that have been determined to be direct contact residues, backbone contact residues, or water-mediated contact residues with the target site of the gene of interest to form a library of engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains;

5 vi) expressing the library of engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains;

 vii) screening the library of engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains for binding activity to the target sequence in the gene of interest and the cleavage activity for the
10 target sequence in the gene of interest; and

 viii) selecting the engineered I-OnuI endonuclease, I-OnuI endonuclease homologues, and hybrids of two I-OnuI or I-OnuI homologue domains that can act upon a nucleotide sequence containing a modification in the target site from the wild-type and directed to a target site within the gene of interest, wherein the binding and cleavage activity
15 is highest for the target sequence in the gene of interest.

5. The methods of any of claims 1 through 4 wherein the I-OnuI homologue comprises about 25% or greater amino acid sequence identity extending over at least 200 amino acids and including both LAGLIDADG sequence motifs with the amino acid sequence of I-OnuI (SEQ ID NO: 35).

20 6. The method of claim 5, wherein the I-OnuI homologue comprises an amino acid sequence that is highly conserved when compared to the LAGLIDADG motifs (amino acid residues 12 to 24 of SEQ ID NO: 35 and amino acid residues 170 to 181 of the I-OnuI amino acid sequence in SEQ ID NO: 35).

7. The method of claim 6, wherein the I-OnuI homologue further comprises high amino acid conservation within a "Loop" sequence adjacent to the first LAGLIDADG helix corresponding to amino acid residues 97 to 103 of SEQ ID NO: 35.

8. The method of claim 7, wherein the I-OnuI homologue demonstrates an overall spacing between the end and beginning of the two LAGLIDADG motifs of between about 162 and 182 amino acid residues.

9. The method of claim 5, wherein the homologues of I-OnuI is I-AabI (SEQ ID NO: 52), I-AaeI (SEQ ID NO: 53), I-ApaI (SEQ ID NO:54), I-CkaI (SEQ ID NO:55), I-CpaI (SEQ ID NO: 56), I-CapIII (SEQ ID NO:57), I-CapIV (SEQ ID NO:58), I-CpaV (SEQ ID NO:59), I-CraI (SEQ ID NO:60), I-EjeI (SEQ ID NO:61), I-GpeI (SEQ ID NO:61), I-GillI (SEQ ID NO:63), I-GzeI (SEQ ID NO:64), I-GzeII (SEQ ID NO:65), I-GzeIII (SEQ ID NO:66), I-HjeII (SEQ ID NO: 67), I-LtrI (SEQ ID NO:68), I-LtrII ,(SEQ ID NO:69) I-MpeI (SEQ ID NO:70), I-MveI (SEQ ID NO:71), I-NcrI (SEQ ID NO:72), I-NcrII (SEQ ID NO:73), I-OheI (SEQ ID NO:74), I-OsoI (SEQ ID NO:75), I-OsoII (SEQ ID NO:76), I-OsoIII (SEQ ID NO:77), I-OsiIV (SEQ ID NO:78), I-PanI (SEQ ID NO:79), I-PanII (SEQ ID NO:80), I-PanIII (SEQ ID NO:81), I-PnoI (SEQ ID NO:82), I-ScuI (SEQ ID NO:83), I-SmaI (SEQ ID NO:84), or I-SscI (SEQ ID NO:85).

1/37

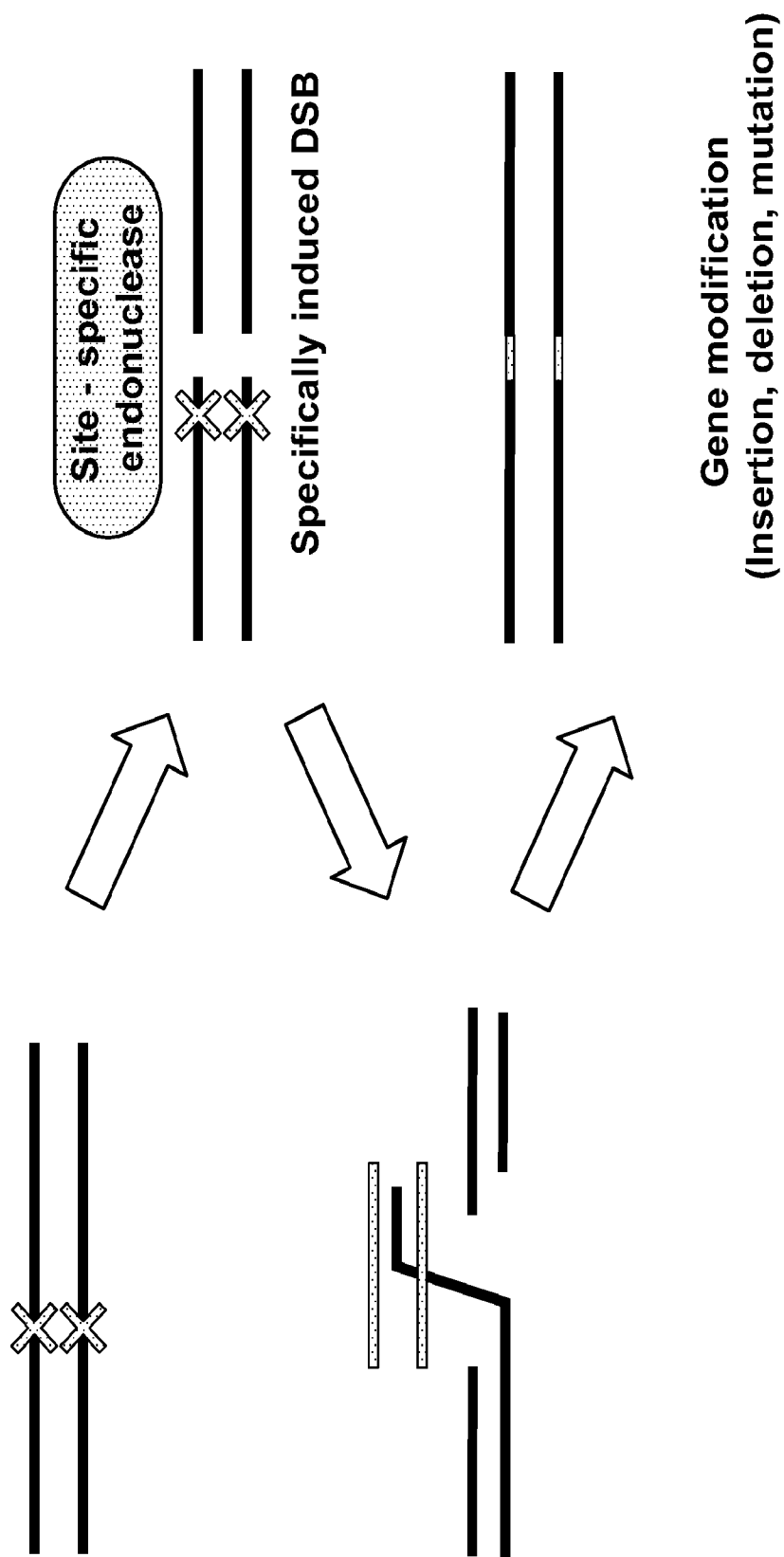


Fig.1.

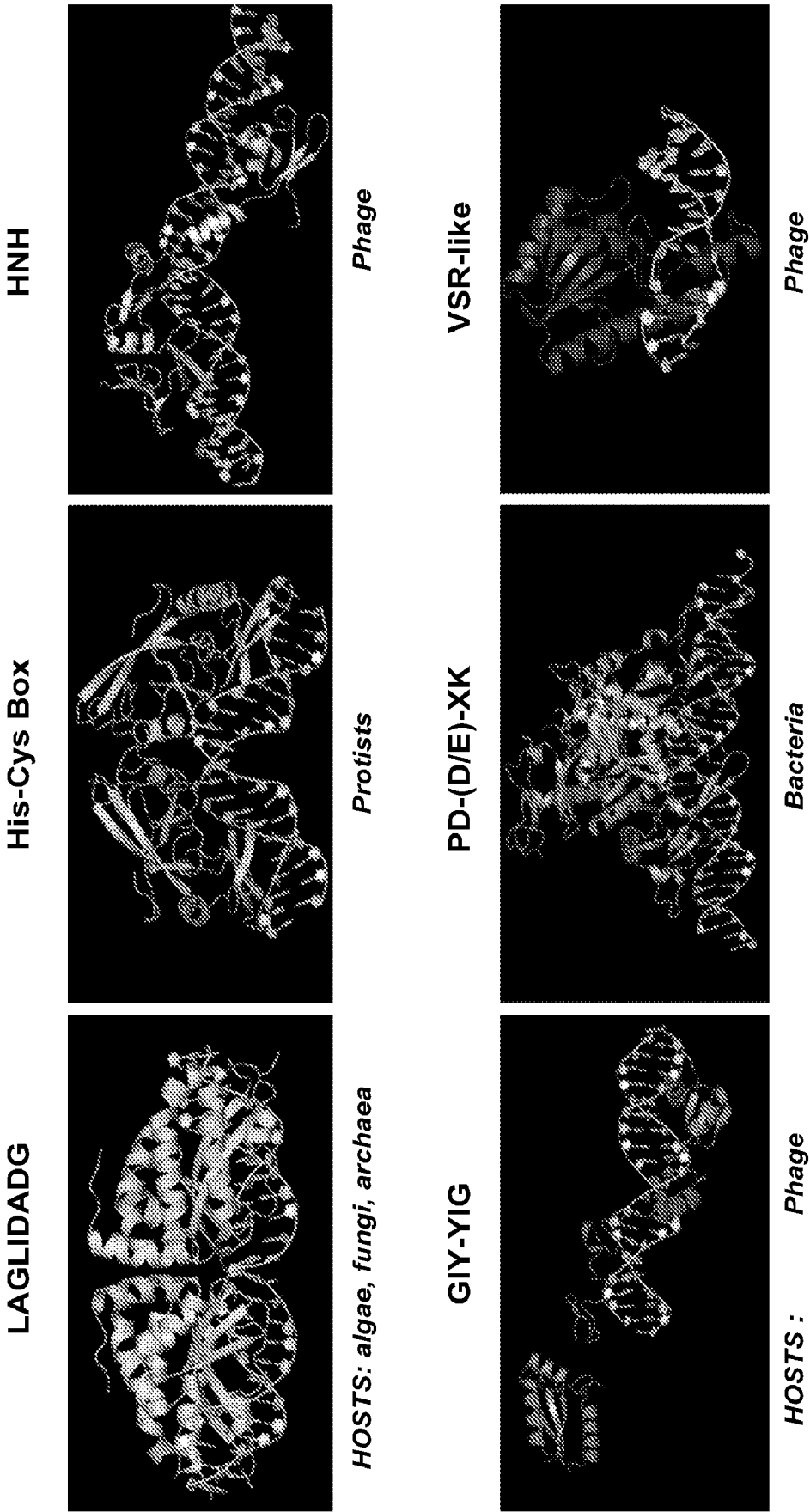


Fig.2.

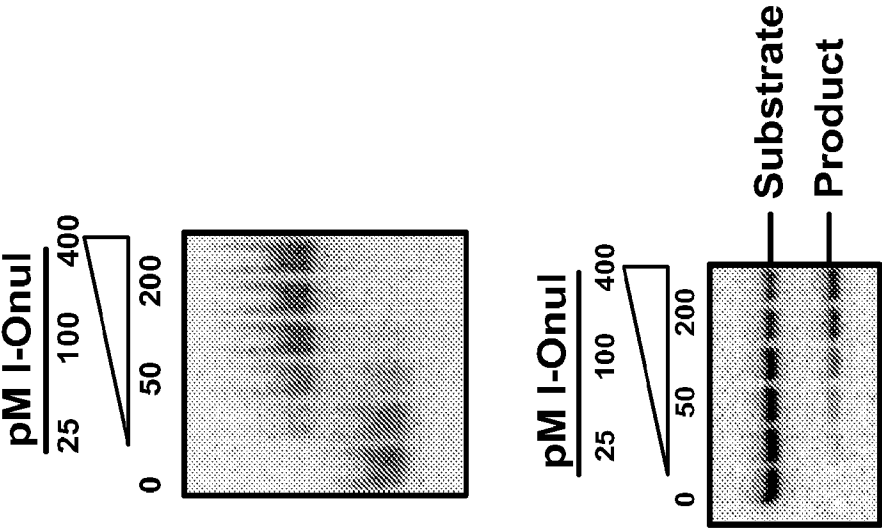


Fig.3B.

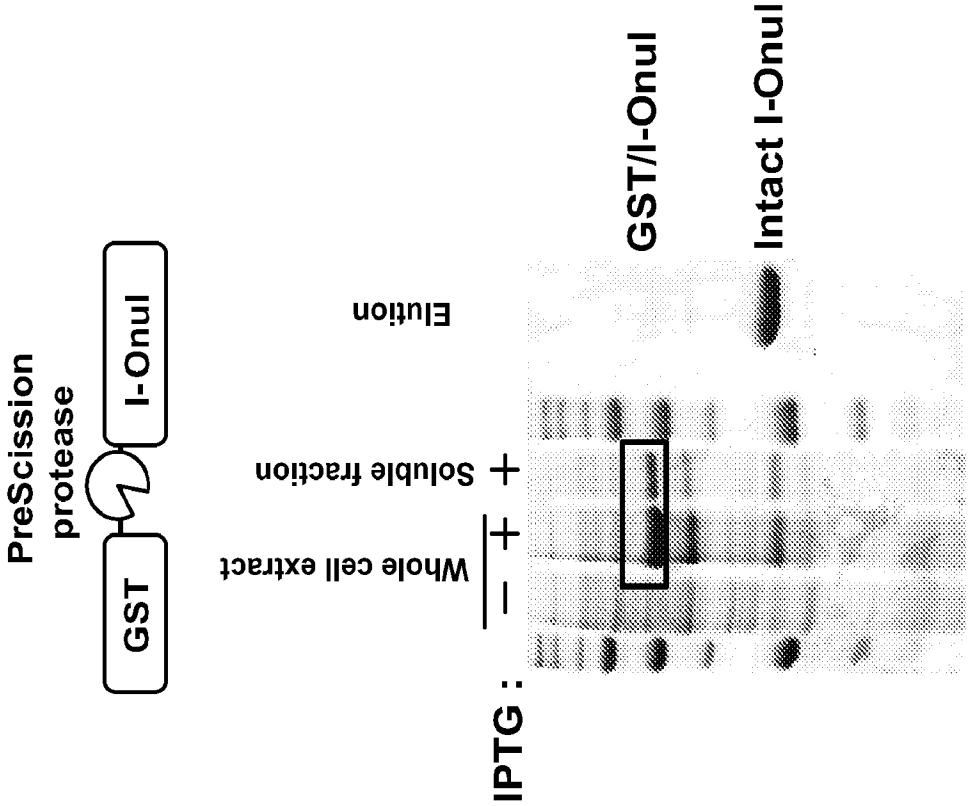


Fig.3A.

Crystal structure of l-Onul bound to the DNA target site

Crystal	Native
<u>Data collection</u>	
Beamline	ALS 8.2.1
Wavelength (Å)	1.000
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell dimension (Å)	a=37.948, b=73.933, c=166.930
Total reflections	124,406
Unique reflections	18,568
Resolution (Å)	50.0-2.40 (2.49-2.40)
Completeness (%)	96.8 (91.0)
Redundancy	6.7 (4.8)
Rmerge	0.080 (0.143)
Average I / σ I	18.6 (9.62)
<u>Refinement</u>	
Rwork (%)	18.9
Rfree (%)	24.0
Rmsd bond length (Å)	0.018
Rmsd bond angle (°)	2.217
Average B factor (Å ²)	13.8
Protein	13.9
Nucleic acid	13.7
Water	11.3
Metal ion	42.7
Ramachandran distribution	
Most favored (%)	90.4
Additionally allowed (%)	8.5
Generously allowed (%)	0.0
Disallowed (%)	0.3

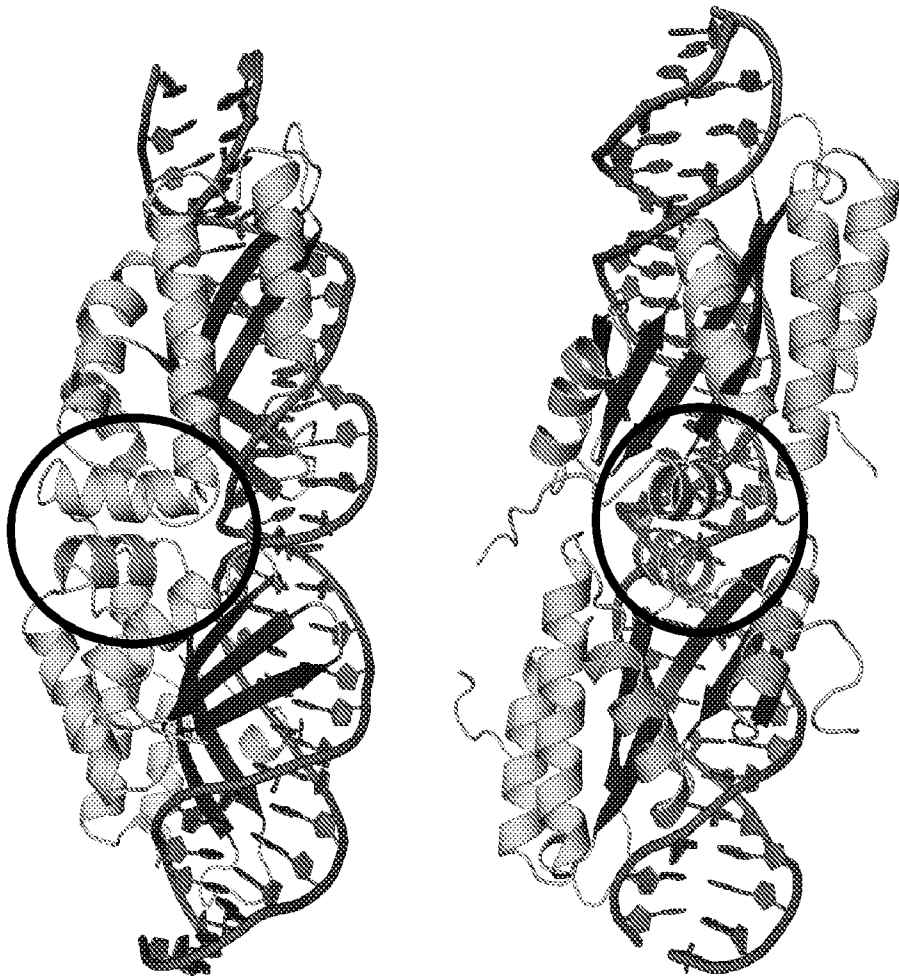
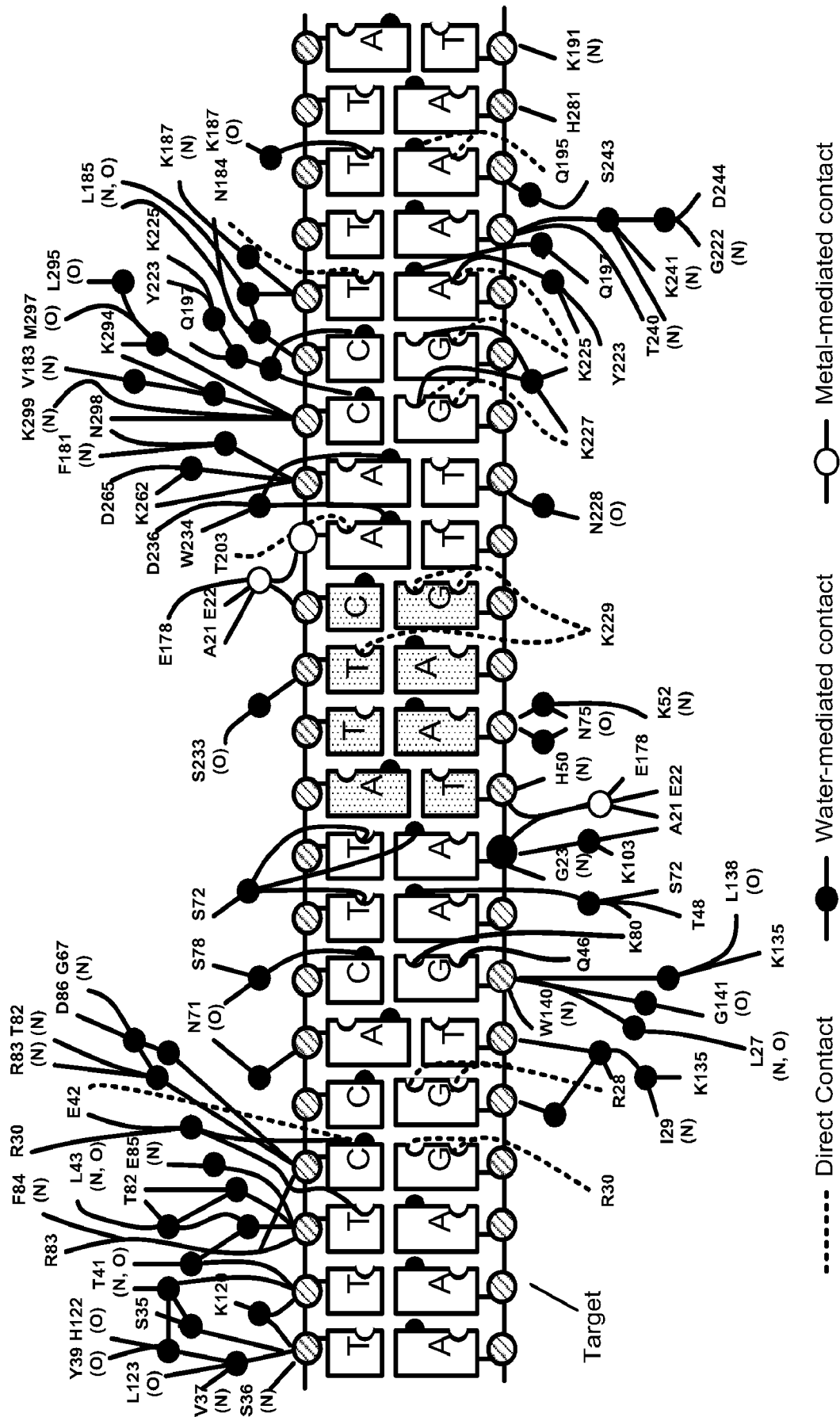


Fig.4.

5/37

*Fig.5.*

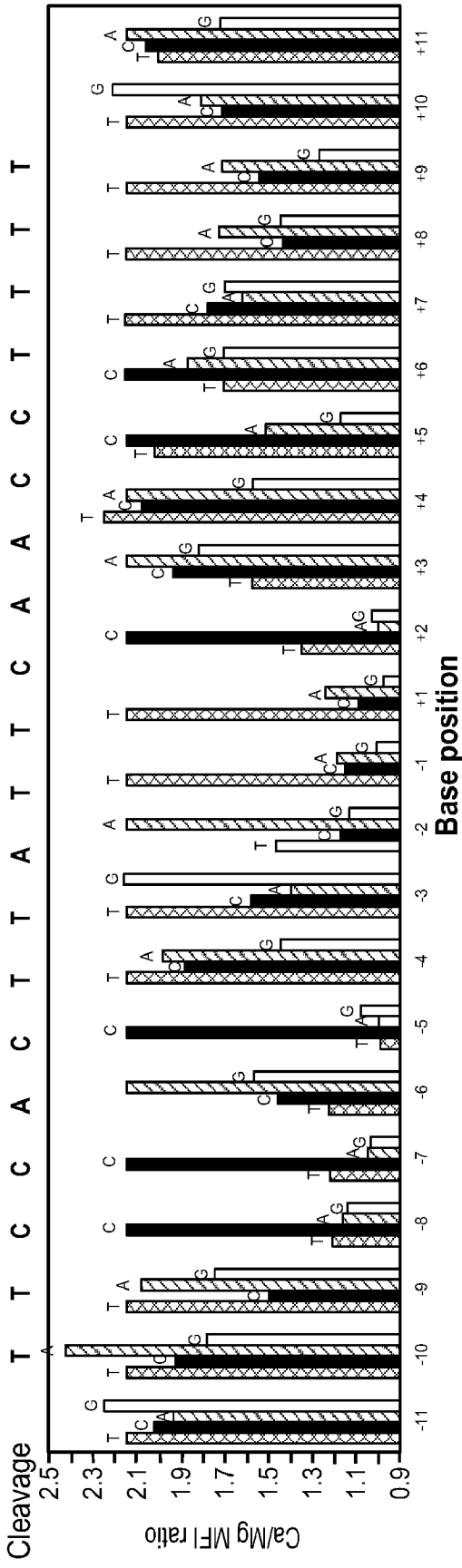


Fig. 6A.

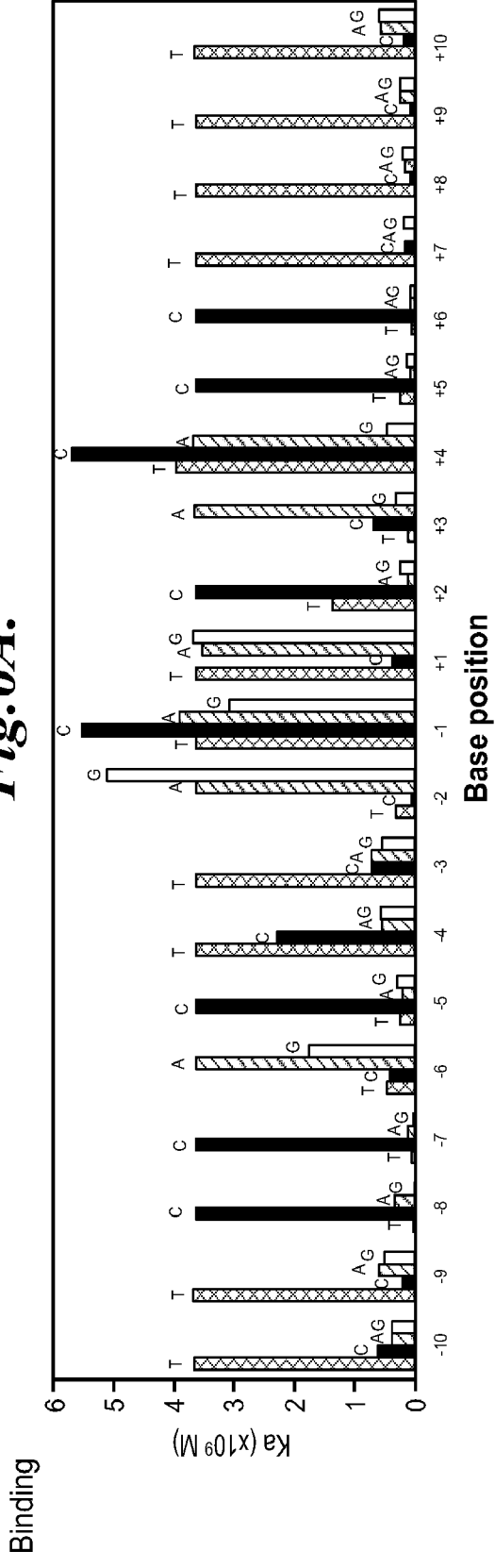


Fig. 6B.

7/37

Base change	Target site	Resultant mutations to WT I Onu-I
-10A	T A TCCACTTATTCAACCTTTTA	S40E
-10C	T C TCCACTTATTCAACCTTTTA	
-10G	T G TCCACTTATTCAACCTTTTA	N32R
-8T	TTT T CACTTATTCAACCTTTTA	<i>R30C/E42L/T82K/R83V/L87F/I90M</i>
-8G	TTT G CACTTATTCAACCTTTTA	
-8A	TTT A CACTTATTCAACCTTTTA	E42Q
-7G	TTT C GACTTATTCAACCTTTTA	K80R
-7A	TTT C A A CTTATTCAACCTTTTA	
-5T	TTTCCATTTATTCAACCTTTTA	S78Q
-5G	TTTCC A GTTATTCAACCTTTTA	
-5A	TTTCC A ATTATTCAACCTTTTA	
-4G	TTTCCAC G TATTCAACCTTTTA	
-4C	TTTCCAC C TATTCAACCTTTTA	
-4A	TTTCCAC A TATTCAACCTTTTA	
-3G	TTTCCACT G ATTCAACCTTTTA	
-3C	TTTCCACT C ATTCAACCTTTTA	T48K
-3A	TTTCCACT A ATTCAACCTTTTA	G73E
-2T	TTTCCACTTTTCAACCTTTTA	<i>S72A/N75R/A76L</i>
+4G	TTTCCACTTATTCA G CTTTTA	
+4T	TTTCCACTTATTCA T CCTTTTA	
+5G	TTTCCACTTATTCA G CTTTTA	
+5T	TTTCCACTTATTCA A TCTTTTA	V238R
+5A	TTTCCACTTATTCA A ACTTTTA	<i>S201Q/K227G/D236V</i>
+6A	TTTCCACTTATTCAAC A TTTA	
+6G	TTTCCACTTATTCAAC G TTTA	V199R
+6T	TTTCCACTTATTCAAC T TTTA	<i>N184A/V199R/K225H</i>
+7A	TTTCCACTTATTCAAC C ATTTA	
+7C	TTTCCACTTATTCAAC C CTTTA	
+8A	TTTCCACTTATTCAAC C ATTTA	<i>N184T/T186Q/Q197R/V199E</i>
+8C	TTTCCACTTATTCAAC C CTTA	
+8G	TTTCCACTTATTCAAC C GTTA	
	computationally designed residues in bold	
	genetically selected residues in <i>italics</i>	

Fig. 7.

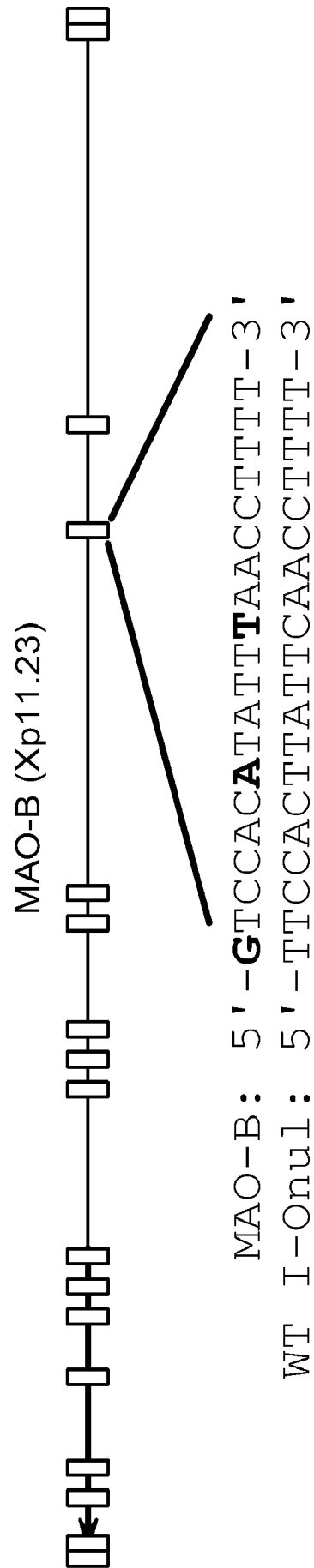
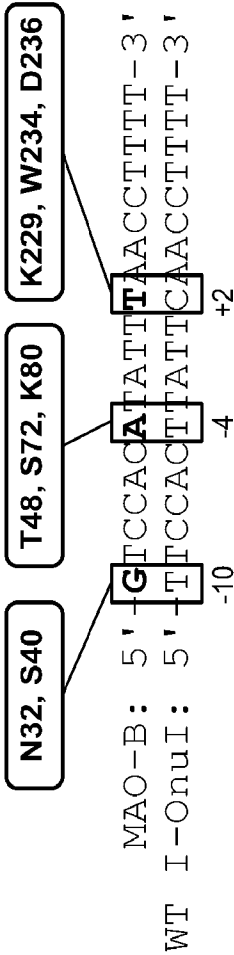


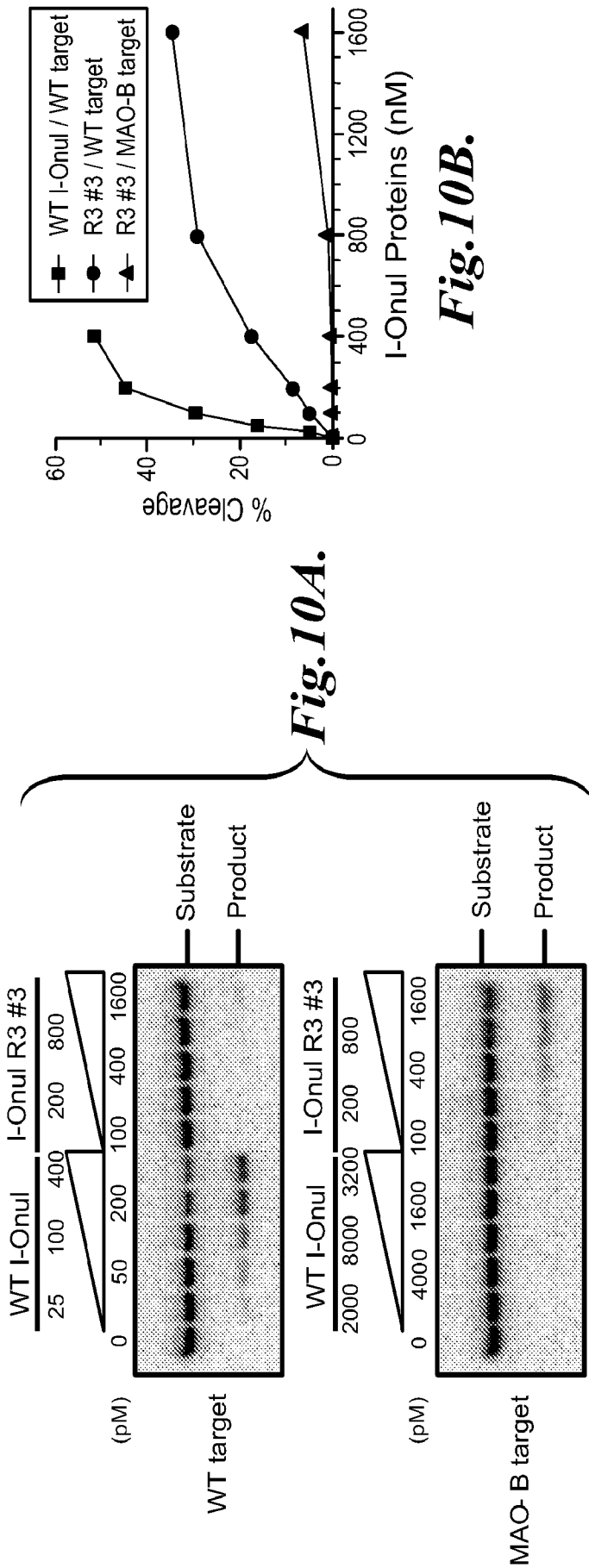
Fig.8.

Summary of directed evolution and engineering of I-Onu1 to redirect specificity towards human mao-B gene



Round	Randomized residues	Pre-induction	Negative selection	Isolated mutants
1	T48, S72, K80	18 hrs	No	T48C/K80R T48C/S72A/K80R T48C/S72C/K80R T48C/S72N/K80R T48C/S72V/K80R T48M/S72R/K80R
2	N32, S40	18 hrs	Yes	N32L/S40R/T48M/S72R/K80R N32L/S40R/T48M/S72A/K80R N32R/S40R/T48M/S72R/K80R N32I/S40R/T48M/S72R/K80R N32L/S40R/T48M/S72R/K80R N32L/S40R/T48M/S72R/K80R/R133C
3	K229, W234, D236	15 hrs	Yes	N32L/S40R/T48M/S72R/K80R/K229R (R3 #3) N32L/S40R/T48M/S72R/K80R/E178D/K229R N32L/S40R/T48M/S72R/K80R/K229H (R3 #6) N32L/S40R/T48M/S72R/K80R/K229Y (R3 #8)

Fig. 9.



	K_D (WT I-Onul)	K_D (Engineered)	k_{cat}^* (WT I-Onul)	k_{cat}^* (Engineered)
WT target	12.3 +/- 1.3 nM	240 +/- 26 nM	58 +/- 4 min ⁻¹	<< 0.1 min ⁻¹
MAO-B target	226 +/- 20 nM	57 +/- 5.6 nM	<< 0.1 min ⁻¹	31 +/- 6 min ⁻¹

Fig. 10C.

11/37

Fig. 11A.

	10	20	30	
<i>OnuI</i>	-	-	-	-
<i>AabI</i>	-	-	-	-
<i>AaeI</i>	-	-	-	-
<i>Apal</i>	-	-	-	-
<i>Ckal</i>	-	-	-	-
<i>Cpal</i>	-	-	-	-
<i>CpalII</i>	-	-	-	-
<i>CpalIV</i>	-	-	-	-
<i>CpaV</i>	-	-	-	-
<i>Cral</i>	-	-	-	-
<i>Ejel</i>	-	-	-	-
<i>Gpel</i>	-	-	-	-
<i>Gpil</i>	-	-	-	-
<i>Gzel</i>	-	-	-	-
<i>GzeII</i>	-	-	-	-
<i>GzeIII</i>	-	-	-	-
<i>HjelI</i>	-	-	-	-
<i>LrlI</i>	-	-	-	-
<i>LrlII</i>	-	-	-	-
<i>Mpel</i>	-	-	-	-
<i>Mvel</i>	-	-	-	-
<i>Ncrl</i>	-	-	-	-
<i>NcrlII</i>	-	-	-	-
<i>Ohel</i>	-	-	-	-
<i>Osol</i>	-	-	-	-
<i>OsolII</i>	-	-	-	-
<i>OsolIII</i>	-	-	-	-
<i>OsolIV</i>	-	-	-	-
<i>PanI</i>	-	-	-	-
<i>PanII</i>	-	-	-	-
<i>PanIII</i>	-	-	-	-
<i>PnoII</i>	-	-	-	-
<i>Scul</i>	-	-	-	-
<i>SmaI</i>	-	-	-	-
<i>Sscl</i>	-	-	-	-

12/37

Fig. 11B.

40	Onul	N	K	S	S	V	G	Y	S	-	T	E	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	Aabl	N	R	-	A	I	D	W	G	-	V	Q	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	Aael	E	N	S	K	L	K	W	G	-	I	Y	P	S	F	A	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	Apal	K	Q	L	R	T	G	W	C	-	V	K	P	V	F	S	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	Ckal	D	N	-	L	I	G	W	R	-	V	Q	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	Cpal	N	N	S	K	L	G	W	V	-	T	K	L	R	L	E	M	S	M	L	K	K	D	R	F	T	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N					
	CpalIII	S	K	Y	S	T	G	W	R	-	I	K	P	V	F	A	I	G	L	H	N	K	K	D	L	E	L	L	K	R	I	Q	S	Y	L	G	-	V	G	K	I	H	I				
	CpalIV	S	S	K	S	L	G	F	-	V	L	P	Y	F	F	E	I	A	L	N	E	K	D	R	D	L	L	L	T	G	I	Q	D	Y	L	G	I	G	-	N	I	T	F				
	CpalV	D	Q	D	K	T	K	I	R	G	V	N	F	S	F	F	T	I	S	L	H	I	D	D	I	A	V	L	K	F	I	K	D	K	L	E	-	I	G	N	I	T	V	Y			
	Cral	E	S	S	V	T	G	Y	S	-	V	L	I	Y	F	F	E	I	A	L	N	K	K	D	K	Q	L	L	L	E	T	I	K	Q	V	L	G	I	D	K	N	L	Y	Y			
Ejel	S	K	Y	T	T	G	W	R	-	I	K	P	I	G	F	Q	I	T	G	L	H	K	K	D	L	D	L	L	L	E	N	I	Q	A	Y	L	G	-	I	G	K	I	H	I			
Gpel	N	K	S	S	A	G	Y	S	-	T	E	L	G	F	F	Q	I	T	L	H	K	K	D	I	S	I	L	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
Gpil	K	K	I	K	C	G	W	N	-	V	T	R	R	F	Q	I	A	L	S	Q	K	D	L	A	L	L	L	E	N	I	Q	S	Y	F	F	K	D	A	G	N	I	Y	I				
Gzel	N	K	Y	K	L	G	W	R	-	V	A	C	R	F	V	I	S	L	H	K	K	D	L	V	L	L	L	E	N	I	Q	S	N	F	F	N	-	T	G	S	V	I	F	L			
Gzell	-	-	N	K	D	K	W	R	-	V	R	P	T	F	Q	I	T	G	L	D	I	K	D	L	S	L	L	E	N	I	Q	S	I	Y	F	N	-	A	G	K	N	I	N	T	K		
GzellIII	K	K	S	G	T	G	W	K	-	I	H	P	T	F	Q	I	T	G	L	D	I	K	D	L	I	L	L	E	N	I	Q	S	I	F	K	A	Y	F	N	-	A	G	K	I	Y	T	
Hjell	D	N	V	K	S	G	W	V	-	I	K	P	R	F	S	I	S	L	H	N	K	K	D	F	V	L	L	E	N	I	Q	S	I	R	E	Y	L	-	G	A	G	I	T	S			
LtrI	S	K	R	N	T	G	W	S	-	V	L	T	Y	F	F	E	I	G	L	H	N	K	D	Q	T	L	L	E	N	I	Q	S	I	K	T	F	D	L	E	C	N	I	T	S			
LtrII	A	T	N	K	T	G	Y	A	-	V	L	T	Y	F	F	E	I	S	L	H	N	K	D	R	I	L	L	E	N	I	Q	S	I	Q	N	F	F	G	-	F	G	I	H	S			
Mpel	S	K	F	K	T	G	W	H	-	A	Y	L	V	F	S	I	S	L	H	N	K	D	R	D	I	L	L	E	N	I	Q	S	I	K	D	C	F	N	-	C	G	T	I	S	E		
Mvel	K	A	G	K	K	T	-	-	-	A	K	L	R	F	S	I	A	V	M	E	N	D	L	E	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K		
NcrI	P	R	Y	K	I	G	W	N	-	V	Q	A	R	F	Q	I	A	V	M	E	N	D	L	E	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K		
NcrII	D	R	R	L	N	G	W	Q	-	V	K	P	I	F	S	I	S	L	H	N	K	K	D	I	S	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K		
Ohel	N	K	Y	A	T	G	W	R	-	I	K	P	I	F	T	I	G	L	H	N	K	K	D	L	D	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
Osol	E	S	S	I	I	G	Y	S	-	I	L	I	Y	F	F	E	I	A	L	N	K	K	D	L	D	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
OsolII	K	N	Y	A	I	N	W	R	-	V	K	A	I	F	F	A	I	G	L	H	N	K	D	L	D	M	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
OsolIII	E	N	N	K	L	K	W	R	-	V	S	A	Y	F	F	S	I	S	L	H	N	K	K	D	L	P	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
OsolIV	S	G	G	S	T	-	-	Y	-	S	E	A	R	F	F	S	I	S	L	H	N	K	K	D	L	D	T	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K
PanI	S	K	Y	K	T	G	W	R	-	V	V	A	V	F	F	S	I	S	L	H	N	K	K	D	L	F	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
PanII	N	K	Y	K	L	G	W	R	-	V	V	C	R	F	F	S	I	S	L	H	N	K	K	D	L	S	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
PanIII	P	R	N	K	L	G	W	C	-	V	Q	A	F	F	Q	I	A	L	H	N	K	K	D	E	A	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
PnoII	K	V	R	A	L	G	W	R	-	A	E	L	K	F	Q	I	G	L	H	N	K	K	D	L	N	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	
Scul	K	P	N	Q	I	R	-	-	-	L	R	A	I	F	F	I	S	Q	H	I	K	D	E	L	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K		
SmaI	S	K	Y	K	T	G	W	L	-	V	V	Q	A	I	F	Q	I	T	V	D	K	K	D	L	F	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K		
Sscl	T	R	-	A	I	G	W	R	-	V	V	Q	L	F	Q	I	N	L	H	N	K	K	D	I	A	L	L	L	E	N	I	Q	S	I	Q	N	Y	F	D	N	I	G	Y	I	S	K	

50	Onul	N	K	S	S	V	G	Y	S	-	T	E	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	Aabl	N	R	-	A	I	D	W	G	-	V	Q	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	Aael	E	N	S	K	L	K	W	G	-	I	Y	P	S	F	A	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	Apal	K	Q	L	R	T	G	W	C	-	V	K	P	V	F	S	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	Ckal	D	N	-	L	I	G	W	R	-	V	Q	L	F	F	Q	I	N	L	H	N	K	D	K	S	I	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	Cpal	N	N	S	K	L	G	W	V	-	T	K	L	R	L	E	M	S	M	L	K	K	D	R	F	T	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N				
	CpalIII	S	K	Y	S	T	G	W	R	-	I	K	P	V	F	A	I	G	L	H	N	K	K	D	L	E	L	L	K	R	I	Q	S	Y	L	G	-	V	G	K	I	H	I			
	CpalIV	S	S	K	S	L	G	F	-	V	L	P	Y	F	F	E	I	A	L	N	E	K	D	R	D	L	L	L	T	G	I	Q	D	Y	L	G	I	G	-	N	I	T	F			
	CpalV	D	Q	D	K	T	K	I	R	G	V	N	F	S	F	F	T	I	S	L	H	I	D	D	I	A	V	L	K	F	I	K	D	K	L	E	-	I	G	N	I	T	V	Y		
	Cral	E	S	S	V	T	G	Y	S	-	V	L	I	Y	F	F	E	I	A	L	N	K	K	D	K	Q	L	L	L	E	T	I	K	Q	V	L	G	I	D	K	N	L	Y	Y		
Ejel	S	K	Y	T	T	G	W	R	-	I	K	P	I	G	F	Q	I	T	G	L	H	K	K	D	L	D	L	L	L	E	N	I	Q	A	Y	L	G	-	I	G	K	I	H	I		
Gpel	N	K	S	S	A	G	Y	S	-	T	E	L	G	F	F	Q	I	T	L	H	K	K	D	I	S	I	L	L	E	N	I	Q	S	T	W	-	K	V	G	V	I	A	N			
Gpil	K	K	I	K	C	G	W	N	-	V	T	R	R	F	Q	I	A	L	S	Q	K	D	L	A	L	L	L	E	N	I	Q	S	Y	F	F	K	D	A	G	N	I	Y	I			
Gzel	N	K	Y	K	L	G	W	R	-	V	A	C	R	F	V	I	S	L	H	N	K	K	D	L	V	L	L	L	E	N	I	Q	S	N	F	F	N	-	T	G	S	V	I	F	L	
Gzell	-	-	N	K	D	K	W	R	-	V	R	P	T	F	Q	I	T	G	L	D	I	K	D	L	S	L	L	E	N	I	Q	S	I	Y	F	N	-	A	G	K	N	I	N	T	K	
GzellIII	K	K	S	G	T	G	W	K	-	I	H	P	T	F	Q	I	T	G	L	D	I	K	D	L	I	L	L	E	N	I	Q	S	I	F	K	A	Y	F	N	-	A	G	K	I	Y	T
Hjell	D	N	V	K	S	G	W	V	-	I	K	P	R	F	S	I	S	L	H	N	K	K	D	F	V	L	L	E	N	I	Q	S	I	R	E	Y	L	-	G	A	G	I	T	S		
LtrI	S	K	R	N	T	G	W	S	-	V	L	T	Y	F	F	E	I	G	L	H	N	K	D	Q	T	L	L	E	N	I	Q	S	I	K	T	F	D	L	E	C						

13/37

Fig. 11C.

	90										100										110																			
Onul	S	G	-	D	N	A	V	S	L	K	V	T	R	F	E	D	L	K	-	V	I	I	D	H	F	E	K	Y	P	L	I	T	Q	K	L	G	D	Y	M	
Aabl	S	-	G	K	N	L	V	Q	Y	R	I	Q	T	F	D	E	L	T	-	I	L	I	K	H	L	K	E	Y	P	L	V	S	K	K	R	A	D	F	E	
Aael	N	S	-	K	T	T	A	L	F	R	V	D	N	L	K	E	L	Q	-	V	I	I	N	H	F	D	K	Y	P	L	V	S	F	K	A	S	D	Y	I	
Apal	H	G	-	P	T	T	L	Q	V	R	F	E	S	K	N	A	L	W	-	A	V	V	K	H	F	D	K	Y	P	L	I	S	Q	K	Q	A	D	Y	L	
Ckal	S	-	G	K	D	T	L	Q	Y	K	I	Q	S	F	T	E	I	K	-	A	I	L	N	H	L	D	K	Y	P	L	I	S	K	K	R	A	D	Y	E	
Cpal	L	D	-	D	N	N	L	R	F	Y	I	E	S	L	K	D	L	T	-	T	V	I	N	H	F	D	K	Y	P	L	F	T	K	K	Q	E	D	Y	L	
CpalII	H	G	-	K	D	S	I	Q	F	R	I	D	S	P	K	E	L	E	-	V	I	I	N	H	F	E	N	Y	P	L	V	T	A	K	W	A	D	Y	T	
CpalIV	N	S	S	D	N	T	F	K	L	K	V	S	G	L	D	G	L	S	N	I	V	I	R	H	F	Q	Q	Y	P	L	Y	L	S	K	N	V	P	I	L	Q
CpalV	K	K	T	R	A	A	C	V	F	S	V	T	N	Q	E	G	L	N	-	K	L	I	S	I	F	D	K	Y	N	L	N	T	T	K	Y	L	D	Y	L	
Cral	N	E	N	D	N	T	L	K	L	R	I	S	N	L	D	V	L	I	K	N	V	I	P	H	F	N	K	Y	P	L	F	T	Q	K	R	V	D	F	L	
Ejel	H	G	-	I	D	S	I	Q	F	R	V	D	S	L	K	E	L	Q	-	V	V	L	N	H	F	D	N	F	P	L	V	T	A	K	L	A	D	Y	I	
Gpel	S	G	-	D	N	A	V	S	L	K	V	T	R	F	E	D	L	R	-	V	V	L	N	H	F	E	K	Y	P	L	I	T	Q	K	L	G	D	Y	L	
Gpil	K	S	D	K	V	S	V	D	W	H	V	T	S	V	K	D	L	K	-	I	I	L	D	H	F	D	K	Y	P	L	K	T	E	K	L	A	D	Y	I	
Gzel	M	-	G	K	G	A	A	Q	Y	R	V	E	S	L	T	G	L	S	-	I	I	I	N	H	F	D	R	Y	P	L	N	T	K	K	Q	A	D	Y	M	
Gzell	S	N	K	E	-	-	C	V	Y	K	V	R	S	L	K	D	I	S	-	I	I	I	S	H	F	D	K	Y	N	L	I	T	Q	K	K	A	D	F	E	
GzellII	S	K	-	R	G	I	I	Y	Y	T	I	G	S	T	K	D	L	L	K	H	V	L	P	H	F	D	K	Y	P	L	S	S	L	K	L	K	D	Y	L	
Hjell	Q	G	-	T	D	S	I	Q	Y	R	V	F	S	I	K	D	L	Q	-	L	V	I	D	H	F	D	K	Y	P	L	I	S	Q	K	F	G	D	Y	S	
Ltrl	-	-	-	D	K	D	A	R	I	R	F	E	S	L	K	E	L	E	-	V	V	I	N	H	F	D	K	Y	P	L	I	T	Q	K	R	A	D	Y	L	
LtrlII	N	P	S	D	D	T	L	K	F	K	V	S	N	I	E	Q	I	V	N	K	I	I	P	F	F	E	K	Y	T	L	F	S	Q	K	R	G	D	F	I	
Mpel	H	G	-	S	N	S	L	K	Y	T	V	K	S	L	N	E	L	Q	-	I	I	I	D	H	F	D	K	Y	L	L	I	T	S	K	L	N	D	Y	K	
Mvel	S	R	T	N	G	M	R	Y	F	T	V	S	K	I	S	D	I	N	N	I	I	I	P	H	F	Q	S	Y	L	L	R	G	T	K	L	L	D	F	E	
Ncrl	I	N	D	R	S	T	V	E	F	R	V	S	D	I	T	S	L	N	N	I	I	I	P	H	F	E	K	Y	Q	L	I	T	N	K	Y	G	D	L	V	
NcrlII	H	G	-	I	D	S	I	Q	Y	R	V	S	S	L	K	N	L	Q	-	I	I	T	D	H	F	D	S	Y	P	L	I	T	Q	K	R	V	D	Y	L	
Ohel	H	G	-	K	D	S	I	Q	F	R	V	D	S	L	K	D	L	Q	-	I	L	L	N	H	F	E	N	F	P	L	V	T	A	K	L	A	D	Y	I	
Osol	N	K	R	D	N	T	L	K	L	R	V	S	N	L	Y	V	L	I	N	N	V	I	P	H	F	N	K	-	-	-	F	T	Q	K	R	V	D	F	L	
OsolII	H	S	-	E	H	S	L	Q	Y	R	V	E	S	I	K	D	L	Q	-	V	I	I	D	H	F	D	N	Y	P	L	V	T	C	K	I	I	D	Y	D	
OsolIII	N	N	-	K	N	T	V	L	F	R	V	S	N	M	Q	E	L	Q	-	I	V	I	D	H	F	K	R	Y	P	L	I	S	A	K	Y	S	D	F	L	
OsolIV	H	G	-	E	D	S	I	Q	Y	A	I	T	S	I	E	Q	L	S	T	L	V	I	P	H	F	D	N	Y	P	L	I	S	K	K	H	A	D	Y	L	
PanI	H	G	-	K	G	T	F	S	Y	R	I	E	S	S	E	Q	I	M	K	F	I	I	P	F	F	D	K	Y	P	L	I	T	E	K	L	G	D	Y	L	
PanII	M	-	T	K	D	S	A	Q	Y	R	V	E	S	L	K	G	L	D	-	L	I	I	N	H	F	D	K	Y	P	L	I	T	K	K	Q	A	D	Y	K	
PanIII	R	G	D	K	-	-	C	T	F	I	V	S	S	L	S	Q	I	T	K	V	I	I	P	H	F	D	S	Y	P	L	I	T	N	K	L	A	D	Y	F	
PnoII	A	K	N	R	D	M	V	N	Y	S	I	D	S	I	K	D	L	N	-	N	L	I	L	Y	L	D	K	Y	P	L	L	T	Q	K	A	A	D	F	L	
Scul	N	K	-	-	K	E	I	Q	Y	V	V	N	S	I	N	D	L	I	N	I	I	I	P	H	F	D	K	Y	P	L	K	Y	G	K	Y	T	S	Y	L	L
Smal	S	G	-	N	S	T	F	S	Y	R	I	E	S	S	S	E	Q	L	T	K	I	I	L	P	F	F	D	K	Y	P	L	I	T	E	K	L	G	D	Y	L
Sscl	S	-	G	T	N	L	V	Q	Y	R	I	Q	T	F	D	E	L	S	-	I	L	I	N	H	L	N	D	Y	P	L	V	S	Q	K	W	D	F	E		

14/37

Fig. 11D.

Onul	120	L	F	K	Q	A	F	C	V	M	E	N	K	-	E	H	L	K	I	N	G	-	-	-	-	I	K	E	L	V	R	I	K	A	K	L	-	N	W	G	L	T	160	
Aabl		L	F	N	T	A	H	K	L	I	K	L	N	-	E	H	L	N	K	E	G	-	-	-	-	I	K	L	V	S	L	K	A	S	L	-	N	L	G	L	S			
Aael		I	F	K	K	C	Y	N	L	I	M	Q	K	-	Q	H	L	T	Q	K	G	-	-	-	-	F	E	E	I	L	S	L	K	Y	K	L	-	N	K	G	L	P		
Apal		L	W	K	K	A	V	E	L	F	S	D	K	-	E	H	L	T	L	A	G	-	-	-	-	L	F	E	V	V	A	L	R	A	S	I	-	N	W	G	L	P		
Ckal		L	F	K	N	A	Y	Y	I	I	L	N	K	-	E	H	L	T	K	L	G	-	-	-	-	L	R	K	L	V	A	L	R	A	S	L	-	N	L	G	L	P		
Cpal		L	F	K	Q	V	F	E	L	F	K	N	K	-	N	H	L	T	M	E	G	-	-	-	-	L	R	Q	I	V	A	I	K	A	L	L	R	N	K	G	L	S		
CpalII		L	F	K	K	A	L	D	V	I	L	L	K	-	E	H	L	S	Q	K	G	-	-	-	-	L	K	L	V	G	I	K	A	S	L	-	N	L	G	L	N			
CpalIV		L	F	S	R	V	V	G	I	M	K	S	G	-	D	H	L	N	L	K	G	-	-	-	-	F	Q	E	I	V	N	I	R	A	S	L	-	N	L	G	L	S		
CpalV		D	F	R	K	A	F	L	L	Y	Q	D	K	D	I	N	F	N	K	E	N	L	N	I	L	I	N	Q	I	V	D	L	K	S	R	M	N	S	E	R	T	F		
Cral		L	M	C	K	I	I	E	L	V	E	E	K	-	K	H	L	T	K	E	G	-	-	-	-	L	N	E	I	L	N	I	K	A	A	M	-	N	L	G	L	S		
Ejel		L	F	K	K	A	F	A	I	I	L	L	K	-	E	H	L	S	S	E	G	-	-	-	-	L	S	K	I	V	G	I	K	A	S	L	-	N	L	G	L	N		
Gpel		L	F	K	Q	A	F	S	V	M	E	N	K	-	E	H	L	K	I	E	G	-	-	-	-	I	K	R	L	V	G	I	K	A	N	L	-	N	W	G	L	T		
Gpil		L	F	K	E	V	F	N	I	I	L	T	K	-	Q	H	L	T	V	E	G	-	-	-	-	I	Q	K	I	V	A	I	R	A	S	I	-	N	K	G	L	Y		
Gzel		L	F	K	L	A	Y	N	L	I	I	N	K	-	S	H	L	T	E	K	G	-	-	-	-	L	S	E	L	V	S	L	K	A	V	M	-	N	N	G	L	K		
GzeII		L	F	K	L	I	I	N	K	L	N	S	Q	-	E	H	L	S	Y	E	V	G	A	T	V	-	L	Q	E	I	S	I	R	A	S	M	-	N	L	G	L	S		
GzeIII		V	F	K	R	I	L	L	M	Q	K	G	-	E	H	Q	S	L	S	G	-	-	-	-	L	F	K	I	F	S	E	R	A	N	L	-	N	K	G	L	P			
Hjell		L	F	K	Q	A	Y	L	L	L	I	N	K	-	E	H	L	T	P	E	G	-	-	-	-	L	K	I	I	A	I	R	A	S	I	-	N	N	G	L	S			
LrI		L	F	K	K	A	F	Y	L	I	K	N	K	-	E	H	L	T	E	E	G	-	-	-	-	L	N	Q	I	L	T	L	K	A	S	L	-	N	L	G	L	S		
LrII		L	F	K	K	A	F	Y	E	L	I	K	N	K	-	E	H	L	T	L	N	G	-	-	-	L	M	K	I	L	S	I	K	A	A	M	-	N	L	G	L	S		
Mpel		L	F	C	K	V	V	E	L	I	K	N	K	-	E	N	L	S	I	E	G	-	-	-	-	L	E	K	L	V	A	I	K	S	S	M	-	N	L	G	L	K		
Mvel		L	F	K	L	A	Y	N	L	F	I	N	K	-	E	N	L	T	E	E	G	-	-	-	-	L	E	K	L	Q	L	L	F	D	G	M	-	-	N	R	K	R		
NcrI		I	F	K	Q	I	V	S	L	M	L	E	N	-	K	H	T	T	L	E	G	-	-	-	-	L	K	E	I	L	E	H	R	A	S	L	-	N	W	G	L	S		
NcrII		L	F	K	Q	A	I	A	L	I	K	N	K	-	E	H	L	S	L	E	G	-	-	-	-	L	L	K	L	V	G	I	K	A	T	L	-	-	R	S	S	W		
Ohel		L	F	K	K	A	F	D	I	I	L	L	K	-	E	H	L	S	Q	E	G	-	-	-	-	L	L	K	L	V	G	I	K	A	S	L	-	N	L	G	L	N		
Osol		S	M	C	K	I	I	E	L	V	Q	W	K	-	S	H	L	T	T	E	G	-	-	-	-	L	N	E	I	L	S	I	K	A	T	M	-	N	L	G	L	S		
OsolI		I	F	K	K	A	F	S	T	I	K	K	Q	-	E	H	L	T	F	E	K	G	-	-	-	-	L	K	L	I	G	L	K	S	S	L	-	N	L	G	L	T		
OsolII		L	F	E	Q	C	Y	N	L	I	K	Q	K	-	E	H	L	T	Q	K	G	-	-	-	-	M	E	K	I	L	A	L	K	S	N	L	-	N	K	G	L	S		
OsolIII		L	F	R	K	A	V	L	L	I	R	N	K	-	K	H	L	T	I	E	G	-	-	-	-	L	Q	E	I	V	S	I	K	A	S	M	-	N	K	G	L	S		
OsolIV		L	F	K	K	V	V	E	M	L	L	N	K	-	E	H	L	T	E	T	G	-	-	-	-	L	Y	K	I	V	S	I	K	A	S	I	-	N	K	G	L	S		
PanI		L	F	K	K	A	H	N	L	I	K	N	K	-	S	H	L	T	K	E	G	-	-	-	-	L	L	E	L	V	A	I	K	A	V	I	-	N	N	G	L	N		
PanII		L	F	K	M	A	H	N	L	I	K	N	K	-	S	H	L	T	K	E	G	-	-	-	-	L	N	E	I	V	R	M	K	A	Q	L	-	N	L	G	L	S		
PanIII		L	W	K	G	I	I	E	I	M	K	A	K	-	K	H	L	T	L	E	G	-	-	-	-	L	N	E	I	V	R	M	K	A	Q	L	-	N	L	G	L	S		
PnoII		L	L	K	K	A	V	E	L	V	N	N	K	-	A	H	L	T	L	E	G	-	-	-	-	L	E	K	I	V	N	I	K	A	S	M	-	N	L	G	L	S		
Scul		I	F	K	N	I	I	E	K	M	K	N	K	-	Q	H	L	T	Q	K	G	-	-	-	-	L	I	D	I	V	N	T	Y	I	M	N	P	L	G	K	R			
Smal		L	F	K	K	V	L	E	L	M	G	T	K	-	E	H	L	T	Q	R	G	-	-	-	-	L	E	K	I	V	S	L	K	A	S	I	-	N	K	G	L	S		
Sscl		L	F	K	Q	A	H	E	L	V	K	M	N	-	E	H	L	N	K	E	G	-	-	-	-	L	I	E	K	I	V	S	L	K	A	S	L	-	N	L	G	L	S	

15/37

Fig. 11E.

[illegible]

SUBSTITUTE SHEET (RULE 26)

Fig. 111.

330	Onul	V	F	Q	E	N	T	L	I	G	V	K	L	E	D	F	E	D	W	C	K	V	A	K	L	I	E	E	K	-	K	H	L	T	E	S	G	L	D	E	
	Aabl	F	F	D	K	Y	P	I	L	G	E	K	A	K	D	L	F	D	F	C	S	V	S	D	L	M	K	S	K	-	T	H	L	T	E	I	G	V	A	K	
	Aael	F	F	N	K	Y	P	I	L	G	V	K	H	L	D	F	Y	D	F	K	L	I	S	N	L	I	D	N	K	-	E	H	L	T	Q	D	R	F	K	-	
	Apal	F	I	R	S	F	P	I	Q	G	M	K	I	H	D	F	N	D	W	C	Q	A	I	D	I	V	R	R	K	-	G	H	L	T	P	E	G	L	A	E	
	Ckal	L	F	N	K	Y	P	I	I	G	Q	K	Y	E	D	F	K	D	F	E	K	V	A	I	L	M	E	N	R	-	K	H	L	T	I	A	G	V	D	E	
	Cpal	I	F	K	E	N	K	I	L	G	E	K	F	K	D	F	Q	D	W	C	R	A	E	L	M	K	N	K	-	A	H	L	T	P	S	G	L	E	E		
	CpalIII	F	F	D	K	Y	S	I	Q	G	K	K	S	L	D	Y	I	N	F	K	E	V	A	D	I	I	K	S	K	-	N	H	L	T	S	E	G	F	Q	E	
	CpalIV	F	F	N	K	Y	P	L	L	G	T	K	S	L	N	Y	Q	D	F	V	K	V	F	F	L	M	D	T	K	-	K	H	L	T	Q	D	G	L	A	T	
	CpalV	Y	L	S	T	E	V	F	K	S	K	K	G	Q	D	F	E	D	W	K	L	I	C	F	E	A	V	Y	K	G	-	S	H	K	I	D	E	I	R	E	L
	Cral	Y	W	K	D	Y	P	L	I	G	N	K	S	K	D	C	L	D	F	I	R	V	Y	E	I	M	L	N	K	-	G	H	L	T	E	E	G	L	T	K	
Ejel	F	F	E	K	Y	L	I	Q	G	K	K	S	L	D	F	I	S	F	K	E	V	A	L	I	V	Q	N	K	-	D	H	L	K	S	E	G	F	K	K		
Gpel	V	F	Q	V	N	N	I	I	G	V	K	L	E	D	F	E	D	W	C	K	V	A	K	L	I	V	E	E	K	-	K	H	L	T	E	S	G	L	E	E	
Gpil	F	F	E	S	I	P	L	Q	G	L	K	L	K	N	F	T	D	F	S	K	A	A	D	I	I	K	V	K	-	G	H	L	T	P	K	G	L	D	K		
Gzel	F	Y	D	K	Y	I	H	G	N	K	K	S	L	D	F	K	D	F	S	R	V	V	T	L	M	E	N	K	-	K	H	L	T	K	Q	G	L	D	E		
Gzeli	L	F	N	E	Y	K	I	K	G	V	K	Y	K	D	F	K	D	W	S	L	V	A	K	M	I	E	S	K	-	S	H	L	T	T	N	G	Y	K	E		
Gzelli	L	L	S	K	H	T	L	S	G	V	K	L	L	D	F	K	E	R	F	K	A	S	I	L	I	N	S	K	-	A	H	L	T	S	E	G	I	K	L		
Hjell	F	F	I	K	Y	P	I	L	G	I	K	A	L	D	F	K	D	F	C	S	L	A	E	L	I	I	S	N	K	-	A	H	Y	T	E	E	G	L	N	K	
Ltrl	F	F	N	K	N	H	I	I	G	V	K	L	L	D	Y	R	D	W	C	K	V	V	T	L	I	I	D	N	K	-	E	H	L	T	S	E	G	L	E	K	
LtrII	F	F	N	E	Y	P	L	I	G	Q	K	L	L	K	N	Y	E	D	F	K	L	I	F	D	M	M	K	T	K	-	D	H	L	T	E	E	G	L	S	K	
Mpel	F	F	S	K	Y	P	I	I	G	D	K	L	L	N	F	Q	N	F	I	E	A	S	K	L	I	N	N	K	-	D	H	L	T	I	E	G	L	K	K		
Nvel	F	F	N	K	Y	P	L	H	G	I	K	L	I	N	L	N	K	I	I	K	I	L	E	L	I	N	K	Y	G	I	N	R	S	N	S	Y	T	S	D		
Ncrl	F	F	D	Q	Y	K	I	E	G	S	K	Y	N	D	Y	V	K	F	K	E	A	I	L	I	I	K	N	K	-	E	H	L	T	E	K	G	L	N	K		
NcrlII	F	L	K	E	Y	P	L	L	G	V	K	Q	E	D	F	L	D	F	V	K	K	I	A	K	L	I	E	S	K	-	T	H	L	T	D	E	G	L	D	T	
Ohel	F	F	E	K	Y	L	I	Q	G	K	K	S	L	D	F	I	S	F	K	E	V	A	L	I	I	K	N	K	-	D	H	L	N	L	E	G	F	K	E		
Osol	Y	W	K	N	Y	P	L	I	G	H	K	K	S	K	D	C	L	D	F	I	R	V	Y	E	I	M	L	T	K	-	G	H	L	T	E	E	G	L	T	K	
OsolII	F	F	E	K	Y	P	I	K	G	K	K	S	L	D	F	L	E	F	N	K	V	V	G	L	V	V	K	N	K	-	Q	H	L	T	Q	E	G	L	D	K	
OsolIII	F	F	N	E	Y	P	I	L	G	V	K	K	L	D	Y	E	D	F	K	K	V	A	E	L	V	V	K	N	K	-	E	H	L	N	V	E	G	L	N	K	
OsolIV	F	F	N	D	H	K	I	V	G	V	K	L	L	Q	D	Y	E	D	F	K	K	V	A	E	L	V	K	A	K	-	G	H	L	T	P	L	G	L	D	F	
PanI	F	F	L	Q	Y	N	I	I	G	S	K	N	L	D	F	N	D	W	C	K	I	A	T	L	M	Q	D	K	-	H	H	L	T	T	E	G	L	N	K		
PanII	F	F	I	K	Y	P	L	K	G	N	K	N	L	D	F	T	D	F	C	E	V	V	R	L	M	E	N	K	-	S	H	L	T	K	E	G	L	D	Q		
PanIII	F	F	K	E	Y	L	P	E	G	N	K	I	K	H	D	Y	L	D	W	C	K	I	H	S	L	M	E	N	N	-	A	H	L	T	E	E	G	L	N	S	
PnoII	F	F	D	E	N	P	I	I	G	I	K	L	L	K	Q	F	I	I	F	K	Q	I	V	N	L	L	Y	N	K	-	S	H	L	T	V	E	G	I	N	S	
Scul	L	F	N	E	N	L	F	F	L	D	K	K	L	K	Q	F	I	I	F	K	Q	I	V	N	L	L	Y	N	K	-	Q	H	L	T	K	D	N	K	L	K	
Smal	F	F	H	Q	Y	K	I	I	G	S	K	Y	G	D	Y	L	Q	D	W	C	K	I	A	L	I	M	Q	N	K	-	N	H	L	T	P	E	G	L	N	E	
Sscl	F	F	D	K	Y	P	I	L	G	E	K	S	K	D	Y	L	D	F	R	T	V	S	E	I	M	R	S	K	-	D	H	L	T	E	V	G	V	A	K		

340	Onul	V	F	Q	E	N	T	L	I	G	V	K	L	E	D	F	E	D	W	C	K	V	A	K	L	I	E	E	K	-	K	H	L	T	E	S	G	L	D	E	
	Aabl	F	F	D	K	Y	P	I	L	G	E	K	A	K	D	L	F	D	F	C	S	V	S	D	L	M	K	S	K	-	T	H	L	T	E	I	G	V	A	K	
	Aael	F	F	N	K	Y	P	I	L	G	V	K	H	L	D	F	Y	D	F	K	L	I	S	N	L	I	D	N	K	-	E	H	L	T	Q	D	R	F	K	-	
	Apal	F	I	R	S	F	P	I	Q	G	M	K	I	H	D	F	N	D	W	C	Q	A	I	D	I	V	R	R	K	-	G	H	L	T	P	E	G	L	A	E	
	Ckal	L	F	N	K	Y	P	I	I	G	Q	K	Y	E	D	F	K	D	F	E	K	V	A	I	L	M	E	N	R	-	K	H	L	T	I	A	G	V	D	E	
	Cpal	I	F	K	E	N	K	I	L	G	E	K	F	K	D	F	Q	D	W	C	R	A	E	L	M	K	N	K	-	A	H	L	T	P	S	G	L	E	E		
	CpalIII	F	F	D	K	Y	S	I	Q	G	K	K	S	L	D	Y	I	N	F	K	E	V	A	D	I	I	K	S	K	-	N	H	L	T	S	E	G	F	Q	E	
	CpalIV	F	F	N	K	Y	P	L	L	G	T	K	S	L	N	Y	Q	D	F	V	K	V	F	F	L	M	D	T	K	-	K	H	L	T	Q	D	G	L	A	T	
	CpalV	Y	L	S	T	E	V	F	K	S	K	K	G	Q	D	F	E	D	W	K	L	I	C	F	E	A	V	Y	K	G	-	S	H	K	I	D	E	I	R	E	L
	Cral	Y	W	K	D	Y	P	L	I	G	N	K	S	K	D	C	L	D	F	I	R	V	Y	E	I	M	L	N	K	-	G	H	L	T	E	E	G	L	T	K	
Ejel	F	F	E	K	Y	L	I	Q	G	K	K	S	L	D	F	I	S	F	K	E	V	A	L	I	V	Q	N	K	-	D	H	L	K	S	E	G	F	K	K		
Gpel	V	F	Q	V	N	N	I	I	G	V	K	L	E	D	F	E	D	W	C	K	V	A	K	L	I	V	E	E	K	-	K	H	L	T	E	S	G	L	E	E	
Gpil	F	F	E	S	I	P	L	Q	G	L	K	L	K	N	F	T	D	F	S	K	A	A	D	I	I	K	V	K	-	G	H	L	T	P	K	G	L	D	K		
Gzel	F	Y	D	K	Y	I	H	G	N	K	K	S	L	D	F	K	D	F	S	R	V	V	T	L	M	E	N	K	-	K	H	L	T	K	Q	G	L	D	E		
Gzeli	L	F	N	E	Y	K	I	K	G	V	K	Y	K	D	F	K	D	W	S	L	V	A	K	M	I	E	S	K	-	S	H	L	T	T	N	G	Y	K	E		
Gzelli	L	L	S	K	H	T	L	S	G	V	K	L	L	D	F	K	E	R	F	K	A	S	I	L	I	N	S	K	-	A	H	L	T	S	E	G	I	K	L		
Hjell	F	F	I	K	Y	P	I	L	G	I	K	A	L	D	F	K	D	F	C	S	L	A	E	L	I	I	S	N	K	-	A	H	Y	T	E	E	G	L	N	K	
Ltrl	F	F	N	K	N	H	I	I	G	V	K	L	L	D	Y	R	D	W	C	K	V	V	T	L	I	I	D	N	K	-	E	H	L	T	S	E	G	L	E	K	
LtrII	F	F	N	E	Y	P	L	I	G	Q	K	L	L	K	N	Y	E	D	F	K	L	I	F	D	M	M	K	T	K	-	D	H	L	T	E	E	G	L	S	K	
Mpel	F	F	S	K	Y	P	I	I	G	D	K	L	L	N	F	Q	N	F	I	E	A	S	K	L	I	N	N	K	-	D	H	L	T	I	E	G	L	K	K		
Nvel	F	F	N	K	Y	P	L	H	G	I	K	L	I	N	L	N	K	I	I	K	I	L	E	L	I	N	K	Y	G	I	N	R	S	N	S	Y	T	S	D		
Ncrl	F	F	D	Q	Y	K	I	E	G	S	K	Y	N	D	Y	V	K	F	K	E	A	I	L	I	I	K	N	K	-	E	H	L	T	E	K	G	L	N	K		
NcrlII	F	L	K	E	Y	P	L	L	G	V	K	Q	E	D	F	L	D	F	V	K	K	I	A	K	L	I	E	S	K	-	T	H	L	T	D	E	G	L	D	T	
Ohel	F	F	E	K	Y	L	I	Q	G	K	K	S	L	D	F	I	S	F	K	E	V	A	L	I	I	K	N	K	-	D	H	L	N								

360

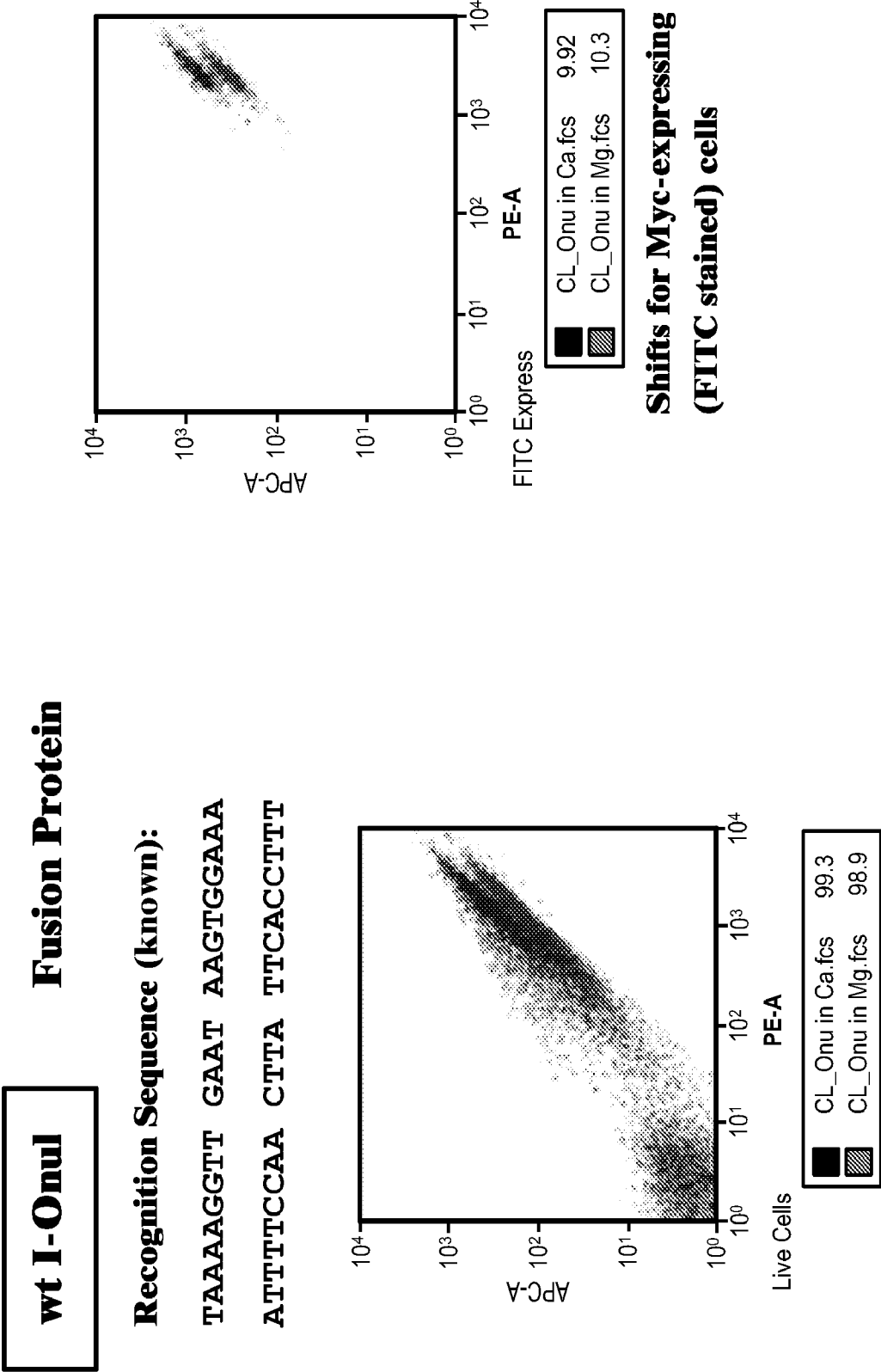
[illegible]

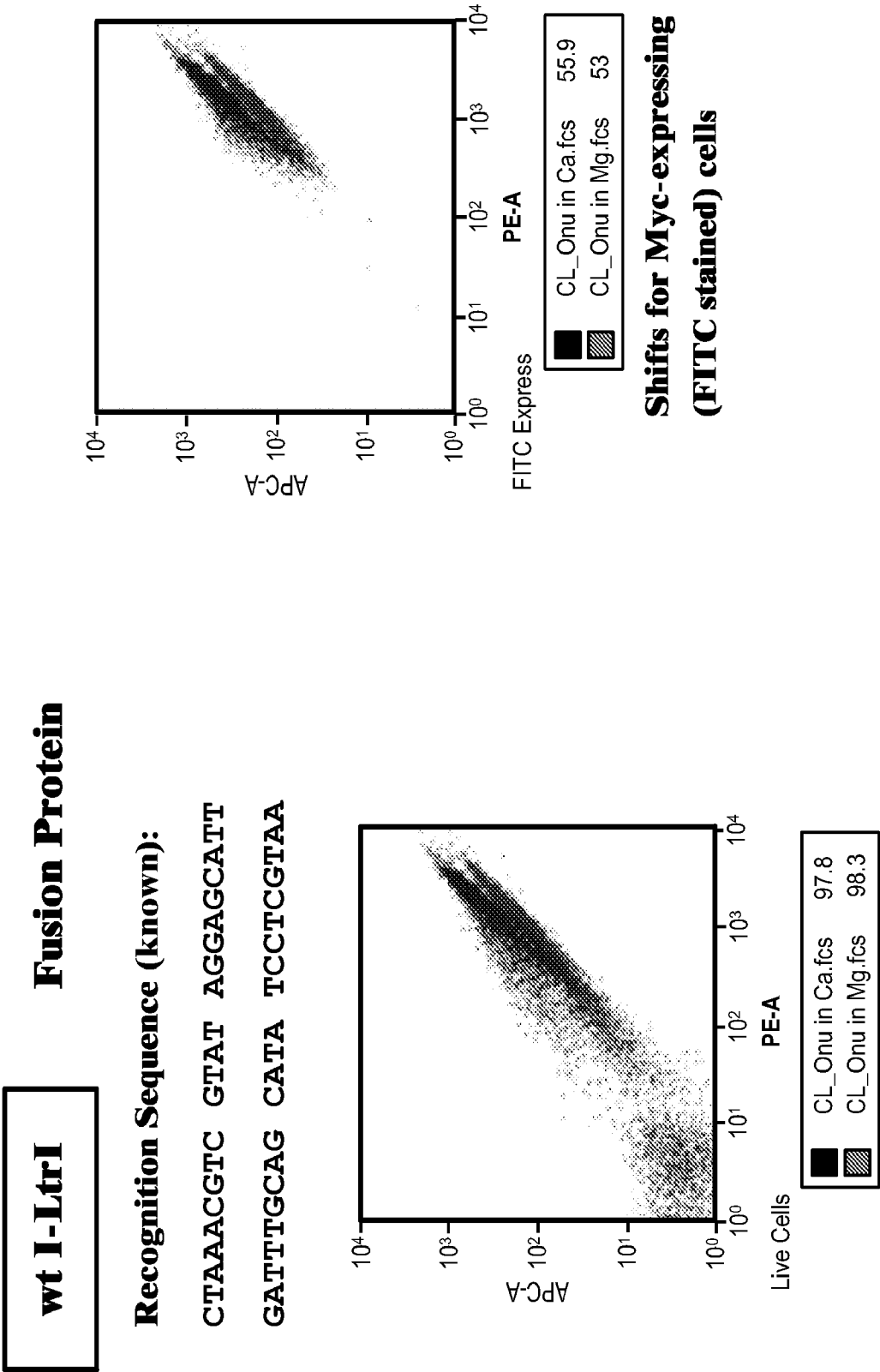
21/37

Crystal structure of I-LtrI bound to its DNA target, shown in two different orientations.



Fig. 12A.





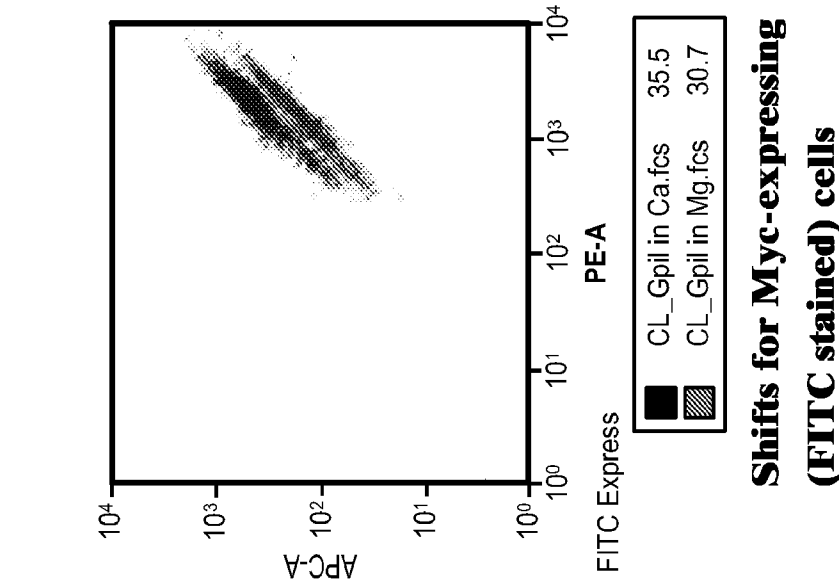
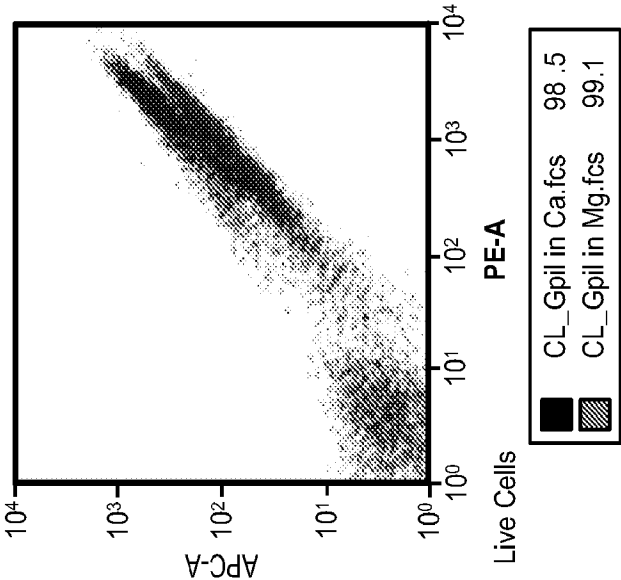


Fig.13C.

I-OnuI	KNOWN	TAAAAAGGTTGAATAAGTGGAAA
I-LtrI	KNOWN	CTAAACGTCGTATAGGAGCATT
Onu LtrI		TAAAAAGGTTGAATAAGTGGAA 22 (Same 12/22) CTAAACGTCGTATAGGAGCATT 22 *** ** * *** * * *
I-GpiI	KNOWN	TTAAAAATTTTCCTGTATATGACTTAAAT
Onu GpiI		-TAAAAAGGTTGAATAAGTGGAA----- 22 (Same 10/22) TTAAAAATTTTCCTGTATATGACTTAAAT 30 ***** ** * * * *
I-MpeI	KNOWN	TTAGCCACTTAIGGTTATCTAAGAGTGTTA
Onu Mpe		-----TAAAAAGGTTGAATAAGTGGAA- 22 (Same 10/22) TTAGCCACTTAIGGTTATCTAAGAGTGTTA 30 * **** * **** *
I-GpeI	PREDICTED	<u>ATAAAGGTTGAATAAGCGGAAAAAT</u>
Onu Gpe		-TAAAAAGGTTGAATAAGTGGAAÑ 22 (Same 20/22) ATAAAGGTTGAATAAGCGGAAAAAT 25 ***** ***** *
I-AabI	PREDICTED	<u>TTATTTCATCTTTTAGTAGGTTTAAAGGGTACCTAGACGGTAT</u>
Onu AabI		-----TAAAAAGGTTGAATAAGTGGAAA----- 22 (Same 12/22) TTATTTCATCTTTTAGTAGGTTTAAAGGGTACCTAGACGGTAT 41 ** ***** ** * *

Fig.14A.

I-SscI	UNKNOWN	- - - -	
I-AaeI	PREDICTED	<u>CCTCATGAAC TGGGCTCTGCAAAAACTTTT</u>	
	Onu	---TAAAAGGTTGAATAAGTGAAA----- 22 (Same 9/22)	
	AaeI	CCTCATGAAC TGGGCTCTGCAAAAACTTTT 30	
		* * * * *	
I-GzeI	PREDICTED	<u>GCCCCCTCATAACCCGTATCAAGAAAAATAT</u>	
	Onu	-----TAAAAGGTTGAATAAGTGAAA--- 22 (Same 10/22)	
	GzeI	GCCCCCTCATAACCCGTATCAAGAAAAATAT 30	
		* * * * *	
I-PanIII	PREDICTED	<u>TTTAGTTGCTTTTCTATAGAT TAAGAGGAAGTATAGCAATAG</u>	
	Onu	-----TAAAAGGTTGAATAAGTGAAA 22	
	PanIII	TTTAGTTGCTTTTCTATAGATTAAGAGGAAGTATAGCAATAG - 41	
		*** * * * *	
I-GzeII	PREDICTED	<u>AAGAA TGACTAAGCATGGGT ACCATATTGGTACAAATTTT</u>	
	Onu	TAAAAGGTTGAATAAGTGAAA ----- 22	
	GzeII	---AAGAA TGACTAAGCATGGGTACCATATTGGTACAAATTTT 40	
		*** * * * *	

Fig.14B.

I-PanI	PREDICTED	<u>CTATTTACCCCTCCTTTATCAGGTGTACAAAAGTCA TA</u>	
Onu		-----TAAAAGGTTGAATAAGTGAAA	22 (Same 12/22)
PanI		CTATTTACCCCTCCTTTATCAGGTGTACAAAAGTCATA	- 36 * * * * *
<hr/>			
I-PanII	PREDICTED	<u>TGCTCCTCATATAATCCT TATCAAGAGAAGTATAGCA</u>	
Onu		-----TAAAAGGTTGAATAAGTGAAA	22 (Same 10/22)
PanII		TGCTCCTCATATAATCCTTATCAAGAGAAGTATAGCA	35 * * * * *
<hr/>			
I-ScuI	PREDICTED	<u>TATGAGGATCCTAAGGT TGTGGGTTAATAT</u>	
Onu		TAAAAGGTTGAATAAGTGAAA	22
ScuI		TATGAGGATCCTAAGGTGTGGGTTAATAT	30 * * * * *
<hr/>			
I-ApaI	PREDICTED	<u>CTCCTGAGGATGCGAAGGT AGCGAAAATTCCTCGAA</u>	
Onu		--TAAAAGGTTGAATAAGTGAAA	22 (Same 10/22)
ApaI		CTCCTGAGGATGCGAAGGTAGCGAAAATTCCTCGAA	35 * * * * *
<hr/>			
I-GzeIII	PREDICTED	<u>AAATTATTGAATATCCT TTAATTATACTATTGTAAAT</u>	
Onu		TAAAAGGTTGAATAAGTGAAA	22
GzeIII		-AAATTATTGAATATCCTTTAAATTATACTATTGTAAAT	38 * * * * *

Fig. 14C.

I-UmaII	PREDICTED	Same as GpeI
I-NcrI	PREDICTED	<u>AACAATCTTATAGGAATGGT TTCTTATAGTTTCGCAAGTAC</u>
Onu		TAAAGGTTGAATAAGTGAAA ----- 22
NcrI		AACAATCTTATAGGAATGGTTTCCTTATAGTTTCG CAAGTAC 42
		* * * * *
I-PnoII	PREDICTED	<u>GCTCCTCATAAATCCT TATCAAGAGAAAGTAT</u>
Onu		-----TAAAGGTTGAATAAGTGAAA N22 (Same 10/22)
PnoII		GCTCCTCATAAATCC TTATCAAGAGAAAGTAT 30
		* * * * *
I-CpaiIII	PREDICTED	<u>AAGTCATCATGGCCT TAAATATTGGGGCTAAAGAC G</u>
Onu		-----TAAAGGTTGAATAA -GTGGAAA----- 22
CpaiIII		AAGTCATCATGGCCTTAAATATT GTGGGCTAAAGACG 36
		* * * * *

Fig.14D.

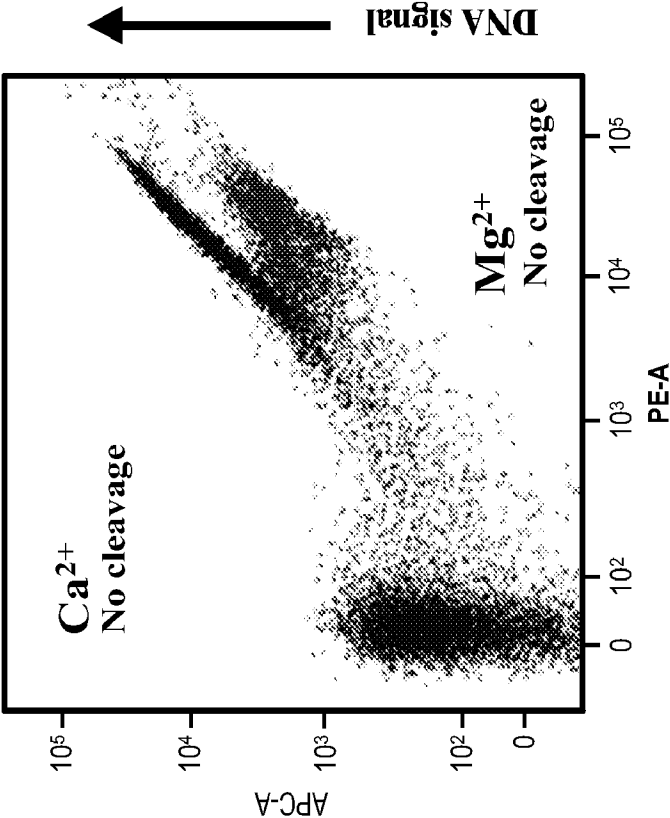
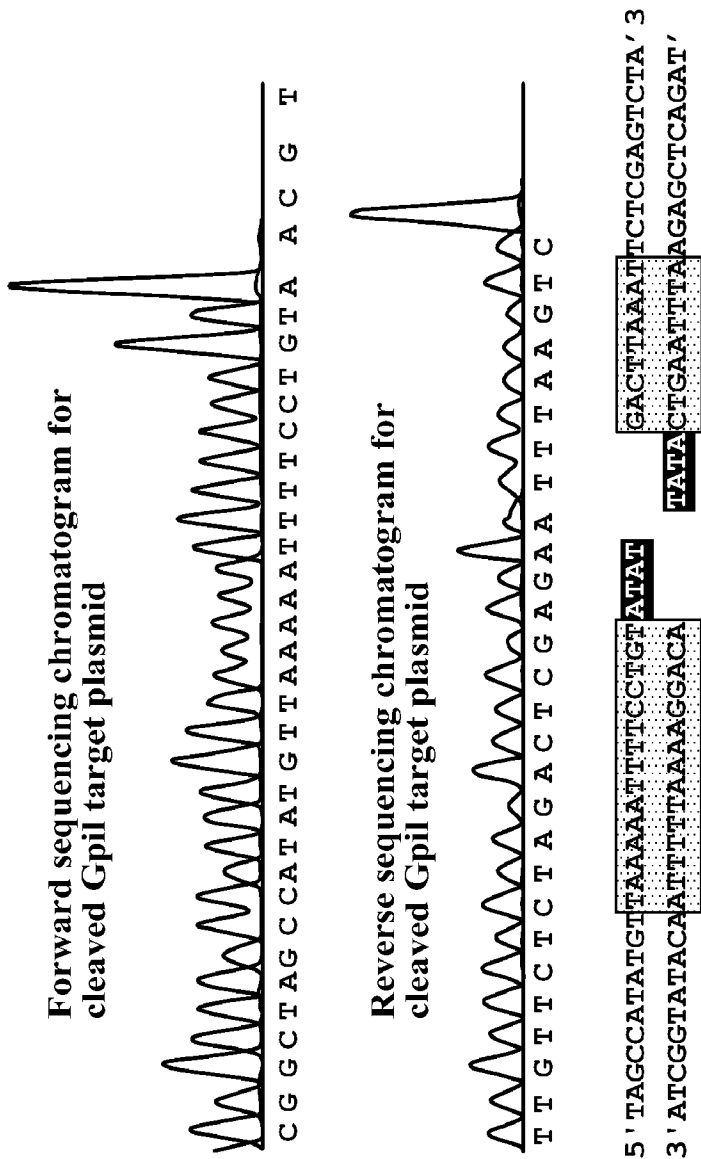


Fig.15B.

I-OnuI	5'-	TTTCCACTTATTCAACCTTTTA	-3'
I-GpeI	5'-	TTTCCGCTTATTCAACCTTTTA	-3'
I-LtrI	5'-	AATGCTCCTATACGACGTTTAG	-3'
I-GpiI	5'-	TTTTCCTGTATATGACTTAAAT	-3'
I-MpeI	5'-	TAGATAACCATAAAGTGGCTAAT	-3'
I-PanII	5'-	GCTCCTCATATAATCCTTATCAAG	-3'
I-GzeI	5'-	GCCCCTCATAACCCGTTATCAAG	-3'
I-Ssci	5'-	AGGTACCCCTTTAAACCTATTAA	-3'
I-Aabi	5'-	AGGTACCCCTTTAAACCTACTAA	-3'
I-PnoII	5'-	CCTTTGGTTATGAGGATCTTCA	-3'
I-GzeII	5'-	TTTGTACCCAAATATGGTACCCAT	-3'
I-CpaiII	5'-	ATGGCCTTAATATTTGTGGGCTA	-3'
I-LtrII	5'-	TTAAATAACATACTTTCACTACT	-3'
I-Smai	5'-	TAATGGAGGATAAATTTGTTTCA	-3'

Fig.15A.



Gpil Target Plasmid Cleavage & Sequencing

Fig.15C.

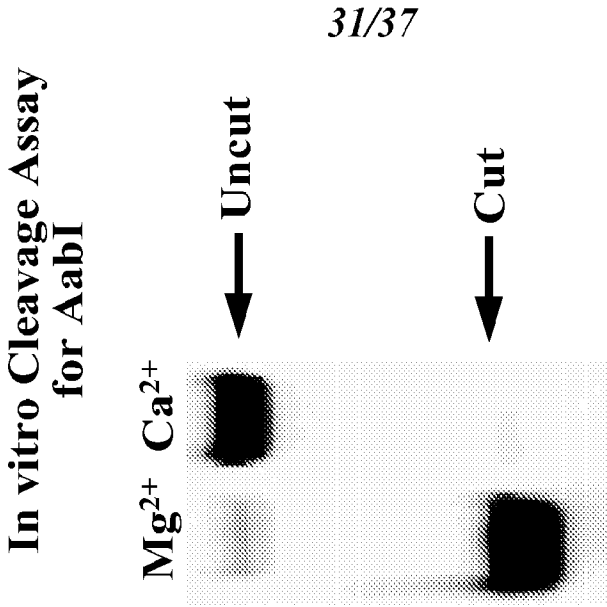
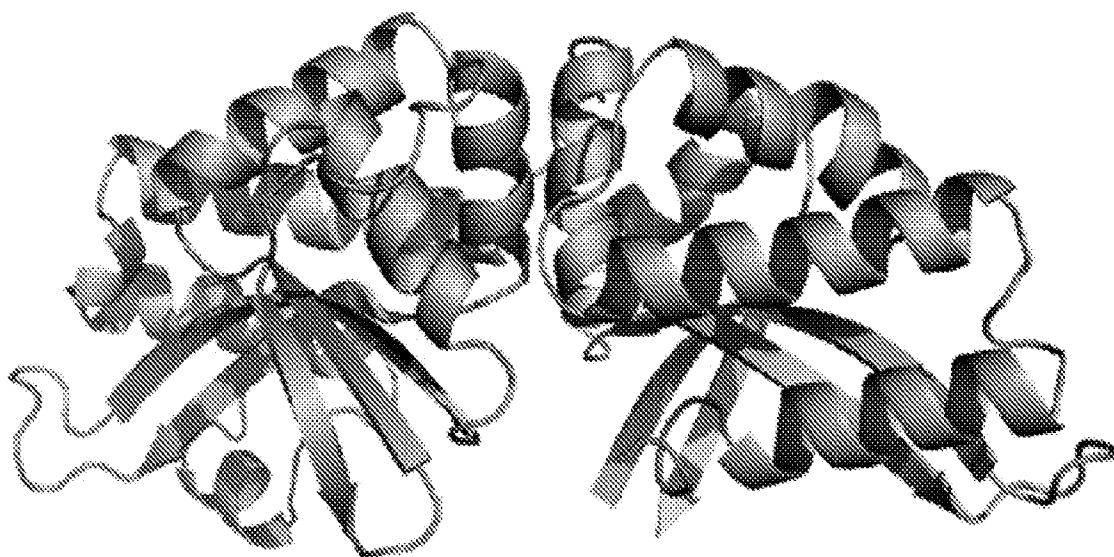


Fig.15D.

32/37



I-Gpil

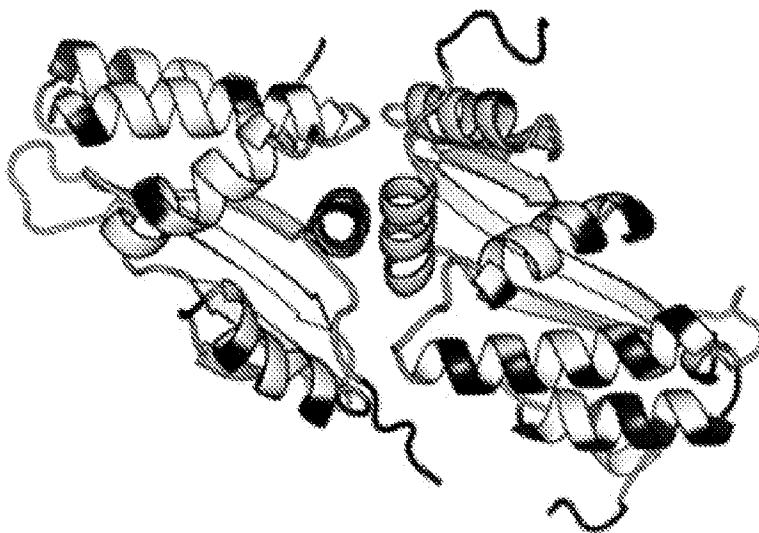
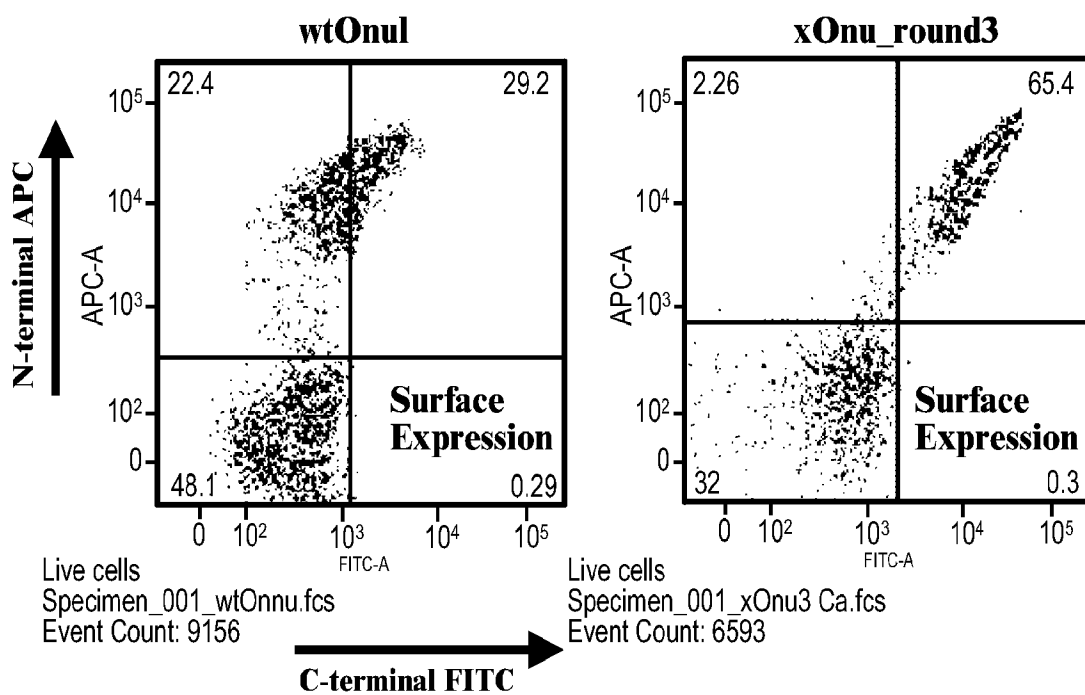


I-Mpel

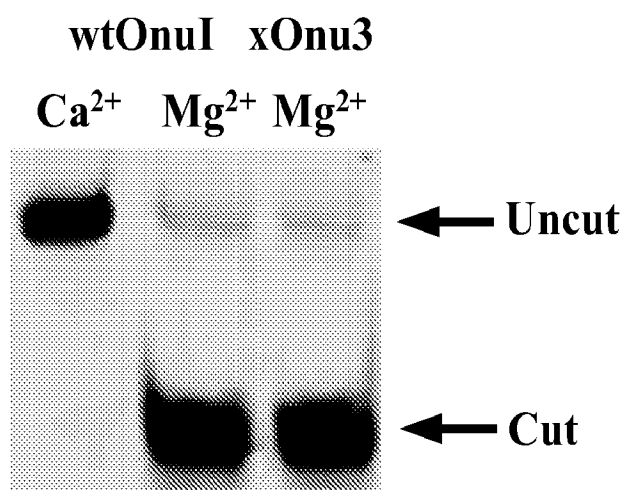
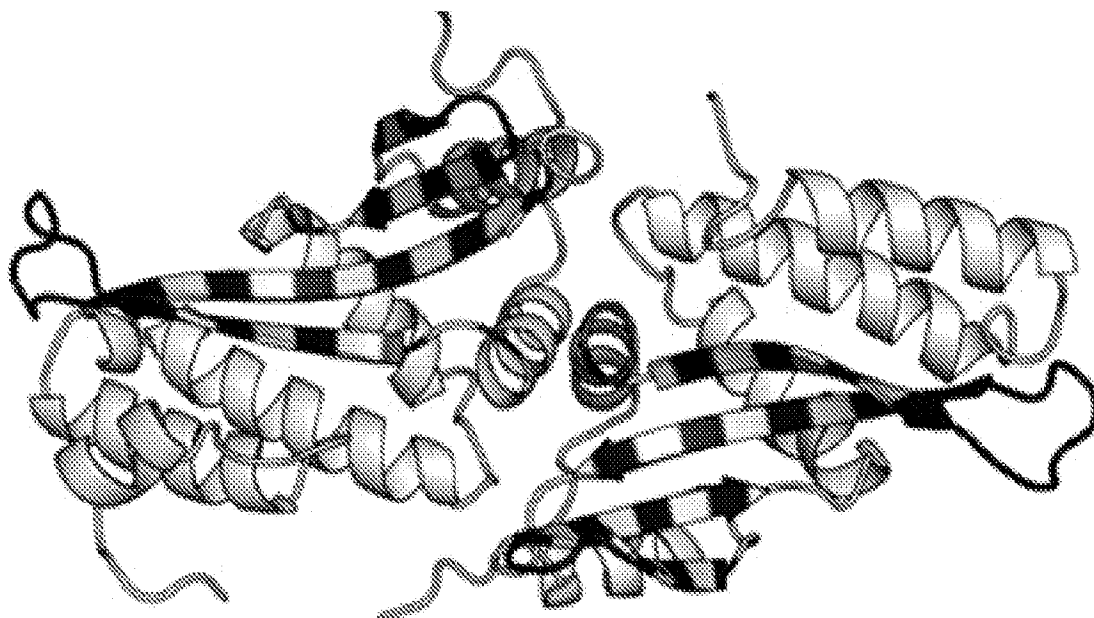
Fig.16.

33/37

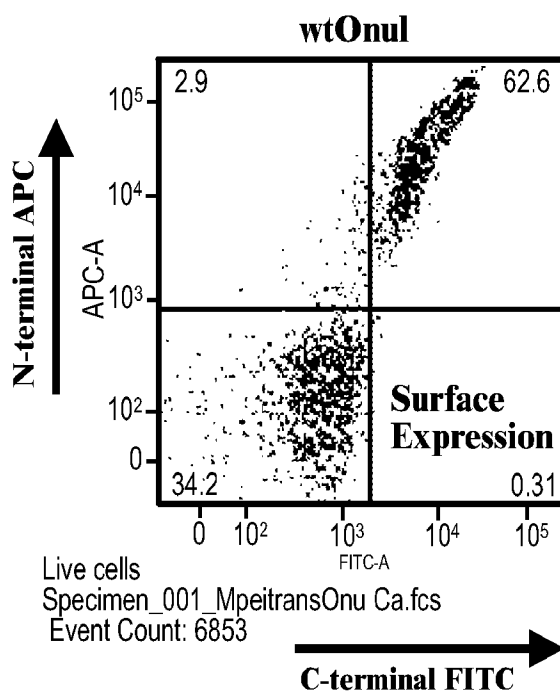
Surface engineering of the solvent-exposed surface

*Fig.17A.**Fig.17B.*

34/37

*Fig.17C.***Surface engineering of the DNA-exposed surface***Fig.17D.*

35/37

*Fig.17E.*

In vitro DNA cleavage assay

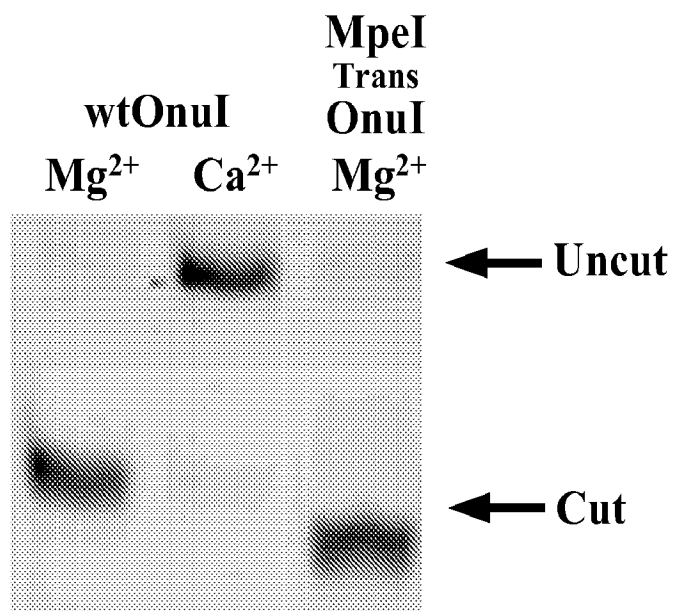
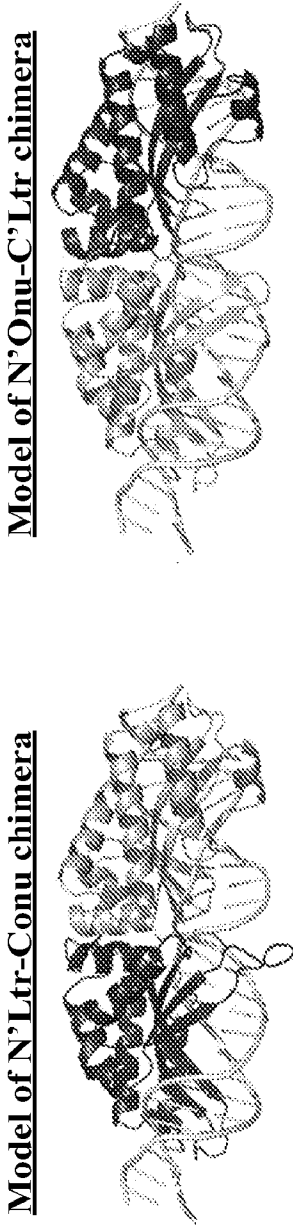
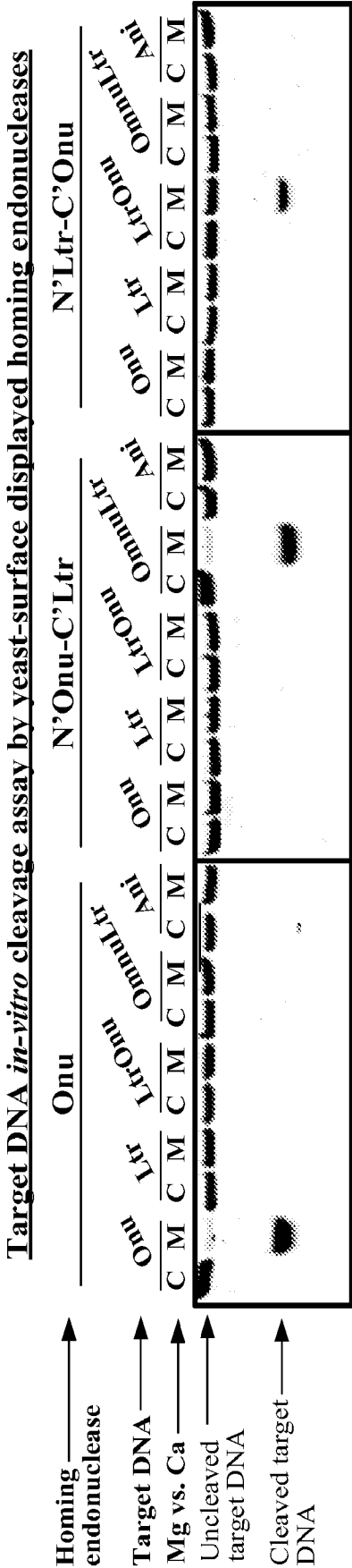
*Fig.17F.*

Fig. 18



Target DNA cleavage by homing endonucleases in flow-cytometry cleavage assay

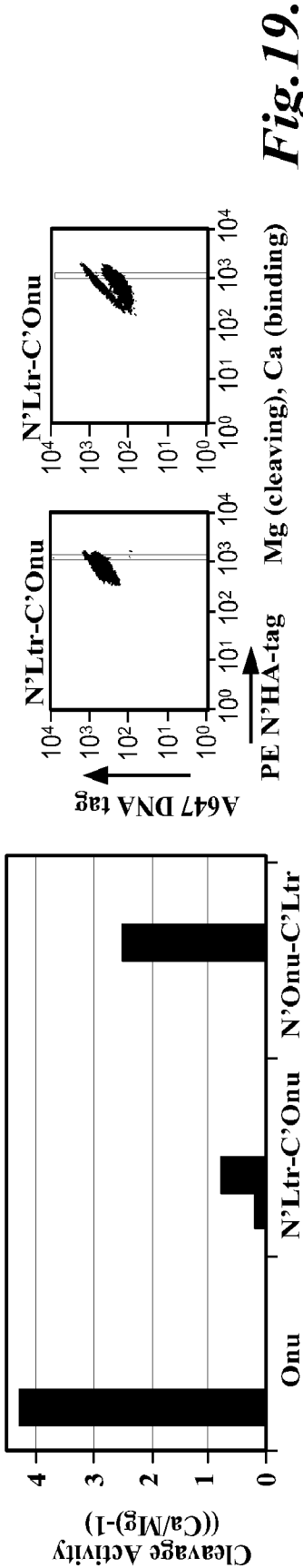


Fig. 19.