(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 January 2012 (19.01.2012)

PCT

(10) International Publication Number
WO 2012/007848 A2

(51) International Patent Classification:
*C07K 14/435* (2006.01)

(21) International Application Number:
PCT/IB2011/002408

(22) International Filing Date:
18 July 2011 (18.07.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/365,096 16 July 2010 (16.07.2010) US

(71) Applicant *(for all designated States except US):* **CELLECTIS** [FR/FR]; 102, Avenue Gaston Roussel, F-93235 Romainville Cedex (FR).

(72) Inventor; and
(75) Inventor/Applicant *(for US only):* **CEDRONE, Frederic** [FR/FR]; 83 Boulevard Diderot, F-75012 Paris (FR).

(74) Agent: **CABINET ORES;** 36 rue de St Petersbourg, F-75008 Paris (FR).

(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, FT, 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:** MEGANUCLEASE VARIANTS CLEAVING A DNA TARGET SEQUENCE IN THE WAS GENE AND USES THEREOF

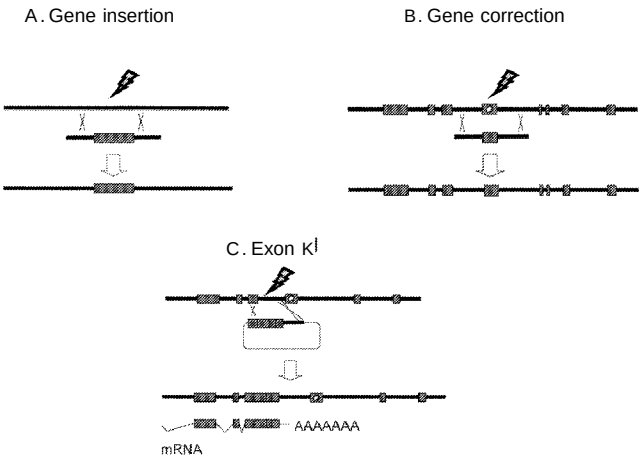


Figure 1

(57) **Abstract:** Meganuclease variants cleaving DNA target sequences of the WAS gene, vectors encoding such variants, and cells expressing them. Methods of using meganuclease variants recognizing WAS gene sequences for modifying the WAS gene sequence or for modulating WAS gene, including gene correction, gene insertion, exon knock-in and gene inactivation.

WO 2012/007848 A2

# TITLE

# MEGANUCLEASE VARIANTS CLEAVING A DNA TARGET SEQUENCE IN THE WAS GENE AND USES THEREOF

## CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Application 61/365,096 filed July 16, 2010, which is hereby incorporated by reference in its entirety.

## BACKGROUND OF THE INVENTION

### Field of the Invention

The present invention concerns a process for manipulating, engineering or modifying polynucleotide sequences of the WAS (Wiscott Aldrich Syndrome) gene locus using endonucleases, such as meganucleases of the I-CreI family. These endonucleases are used in gene insertion, gene correction and exon knock-in ("KI") procedures and are useful in gene therapy of subjects having mutations or defects at this locus.

## DESCRIPTION OF THE RELATED ART

### The Wiskott-Aldrich Syndrome

Wiskott-Aldrich syndrome (WAS) is an X-linked recessive immunodeficiency characterized by thrombocytopenia, eczema, and recurrent infections (Lemahieu, Gastier et al. 1999). The manifestations of Wiskott-Aldrich syndrome are eczema, thrombocytopenia, proneness to infection, and bloody diarrhea. Death usually occurs before age 10 years. The first cases have been described by Aldrich et al. (Aldrich, Steinberg et al. 1954) and Wiskott (Wiskott, A. Familiarer, angeborener Morbus Werlhofii Mschr. Kinderheilk. 68: 212-216, 1937.). More than 160 different WAS mutations spanning all 12 exons of the gene had been found in more than 270 unrelated families and functional domains had been defined (Puck and Candotti 2006). The human WAS can be treated by bone marrow transplantation (Bach, Albertini et al. 1968; Gatti, Meuwissen et al. 1968; Puck and Candotti 2006). However, it could also be treated by gene therapy. A first attempt was initiated by Dr. Christopher Klein in

2007 in Germany, and a second set of clinical trials is being planned by Dr. Adrian Thrasher and his colleagues in three centers in Europe - London, Paris and Milan.

Gene therapy

Gene therapy has been used with success for the treatment of X-SCID patients (Cavazzana-Calvo, Hacein-Bey et al. 2000; Gaspar, Parsley et al. 2004) and significant success have been observed in other clinical trials addressing different diseases, including another form of SCID (Aiuti, Slavin et al. 2002), Epidermolysis Bullosa (De Luca, Pellegrini et al. 2009) and Leber Amaurosis (Bainbridge, Smith et al. 2008; Maguire, Simonelli et al. 2008). However, a series of severe adverse events (SAEs), i.e. the appearance of leukemia, were observed in X-SCID treated patients (Hacein-Bey-Abina, Von Kalle et al. 2003; Hacein-Bey-Abina, Garrigue et al. 2008; Howe, Mansour et al. 2008). Malignant transformations were also observed, in the myeloid lineage of chronic granulomatous disease (CGD) patients treated by gene therapy (Adjali, Marodon et al. 2005 ). In the X-SCID trial, all cases of leukemia, but one, could eventually be treated by chemiotherapy, and the approach appears globally as a success, but these SAEs highlighted the major risks of current gene therapy approaches. Indeed, most of the gene therapy protocols that are being developed these days for the treatment of inherited diseases are based on the complementation of a mutant allele by an additional and functional copy of the disease-causing gene. In non-dividing tissues, such as retina, this copy can be borne by a nonintegrative vector, derived for example, from an Adeno Associated Virus (AAV) (Bainbridge, Smith et al. 2008; Maguire, Simonelli et al. 2008). However, when targeting stem cells, such as hematopoietic stem cells (HSCs), whose fate is to proliferate, persistent expression becomes an issue, and there is a need for integrative vectors. Gamma-retroviral and lentiviral vectors, which integrate in the genome and replicate with the hosts' chromosomes, have proved efficient for this purpose (Chang and Sadelain 2007), but the random nature of their insertion has raised various concerns, all linked with gene expression. The cases of malignant transformation observed in the X-SCID and CGD trials were clearly linked with the activation of proto-oncogenes in the vicinity of the integration sites (Hacein-Bey-Abina, Von Kalle et al. 2003; Adjali, Marodon et al. 2005 ; Hacein-Bey-Abina, Garrigue et al. 2008; Howe, Mansour et al. 2008). Several strategies have been developed to address these different issues. For example, new generations of safer viral vectors, like the Self Inactivating (SIN) gamma-retroviral and lentiviral vectors, should

alleviate the issues concerning activation of nearby potential oncogenes by the viral LTRs (Yu, von Ruden et al. 1986; Yee, Moores et al. 1987; Montini, Cesana et al. 2009)

Targeted approaches for gene therapy

In order to alleviate the drawback of random insertion, a number of targeted approaches have been envisioned (Galetto, Duchateau et al. 2009). Site-specific integration in a safe locus (Fig. 1A) is a promising alternative to random integration of viral vectors. The major challenge is the availability of a region in the genome that could be considered as a "safe harbor" for gene addition. This locus should be chosen in a way that the probability of insertional mutagenesis would be minimized, thus providing for a long-term and high level of expression of a transgene. However, efficient tools to promote site specific and permanent integration in such a chosen locus are needed.

Another approach involves the correction of a mutated gene itself. A gene correction strategy requires very precise genome editing at a targeted locus. Gene correction offers the advantage of directly addressing the cause of a genetic disease: instead of compensating the effect of the mutation by a second genome alteration (such as an insertion in a safe harbor), the true reversion of the disease-causing mutation is the least invasive event one can imagine. However, this precision comes with an inherent drawback: the correction of the mutation, usually based on homologous gene repair, is a very local event, and one needs a different genome engineering tool for each disease, and in most cases, for each mutation or at least each mutation hotspot related to the disease.

Another alternative and intermediate approach between targeted gene correction and gene addition, named here is "exon knock-in". In this approach, a complete or partial cDNA of the affected gene would be integrated in the targeted endogenous locus. This genomic insertion would be less invasive to the cellular genome, since the locus itself would act as a kind of safe harbor for the specific disease. However, this does not alleviate all the possible risks: the resulting gene could lack sequences involved in gene regulation if they are found in the missing introns. Additionally, the genomic locus would be significantly modified, with potential consequences at the transcriptional level.

Sequence specific endonucleases recognizing long target sites appear to be one of the most promising tools for targeted approaches for gene therapy (Galetto, Duchateau et al. 2009). Such endonucleases, including Zinc-Finger Nucleases and Meganucleases can

recognize and cleave a specific sequence in living cells, generating a DNA double strand break (DSB). The outcome could be either re-joining of the DSB by NHEJ, or homologous recombination between the broken locus and one episomal DNA molecule of related sequence (Paques and Duchateau 2007). Zinc Finger Nucleases (ZFNs) are fusions of engineered eukaryotic Zinc Finger DNA binding protein with the catalytic domain of the bacterial Type IIS FokI restriction enzyme (Kim, Cha et al. 1996; Smith, Bibikova et al. 2000). Recently, ZFNs have been used to inactivate the CCR5 gene, a co-receptor for HIV entry, whose inactivation leads to resistance to HIV infection (Perez, Wang et al. 2008). This work was key to the opening, in 2009, of a new clinical trial for AIDS patients.

Meganucleases

Meganucleases are sequence specific endonucleases whose function in nature is to induce targeted recombination (Paques and Haber 1999; Korkegian, Black et al. 2005). Meganucleases were the first endonucleases used to induce DSBs and recombination in living cells (Rouet, Smih et al. 1994; Rouet, Smih et al. 1994; Choulika, Perrin et al. 1995; Puchta, Dujon et al. 1996). They can cleave unique sites in living cells, thereby enhancing gene targeting by 1000-fold or more in the vicinity of the cleavage site (Puchta et al., Nucleic Acids Res., 1993, 21, 5034-5040 ; Rouet et al., Mol. Cell. Biol., 1994, 14, 8096-8106 ; Choulika et al., Mol. Cell. Biol., 1995, 15, 1968-1973; Puchta et al., Proc. Natl. Acad. Sci. U.S.A., 1996, 93, 5055-5060 ; Sargent et al., Mol. Cell. Biol., 1997, 17, 267-277; Cohen-Tannoudji et al., Mol. Cell. Biol., 1998, 18, 1444-1448 ; Donoho, et al., Mol. Cell. Biol., 1998, 18, 4070-4078; Elliott et al, Mol. Cell. Biol., 1998, 18, 93-101).

These proteins can be classified into five families (Chevalier and Stoddard 2001; Stoddard 2005). The largest and best characterized, is the LAGLIDADG family. Structural studies of LAGLIDADG proteins identified a very conserved core structure (Stoddard 2005), characterized by an αββαββα fold, with the LAGLIDADG motif belonging to the first helix of this fold . A number of laboratories have specifically mutagenized residues from the DNA-binding interface of these proteins (Seligman, Chisholm et al. 2002; Gimble, Moure et al. 2003; Arnould, Chames et al. 2006), and characterize the resulting changes in activity and/or specificity. In addition, different assays were developed in order to directly test the activity of meganuclease variants in living cells (Seligman, Chisholm et al. 2002; Gimble, Moure et al. 2003; Chames, Epinat et al. 2005; Chen and Zhao 2005; Arnould, Chames et al. 2006; Doyon,

Pattanayak et al. 2006; Chen, Wen et al. 2009). *In vivo* assays are in principle more relevant if the goal is to develop genome engineering tools as opposed to a mere restriction enzyme for *in vitro* use. Such assays were based either on the inactivation of a plasmid in bacteria (Chen and Zhao 2005; Doyon, Pattanayak et al. 2006; Chen, Wen et al. 2009), or on the induction of homologous recombination in yeast or mammalian cells (Epinat, Arnould et al. 2003; Arnould, Chames et al. 2006; Li, Pellenz et al. 2009; McConnell Smith, Takeuchi et al. 2009). Although several hundred natural meganucleases, also referred to as "homing endonucleases" have been identified (Chevalier, B.S. and B.L. Stoddard, Nucleic Acids Res., 2001, 29, 3757-3774), the repertoire of cleavable target sequences is too limited to allow the specific cleavage of a target site in a gene of interest or GOI as there is usually no cleavable site in a chosen gene of interest.

Theoretically, the making of artificial sequence-specific endonucleases with chosen specificities could alleviate this limit. To overcome this limitation, an approach adopted by a number of workers in this field is the fusion of Zinc-Finger Proteins (ZFPs) with the catalytic domain of Fokl, a class IIS restriction endonuclease, so as to make functional sequence-specific endonucleases (Smith et al., Nucleic Acids Res., 1999, 27, 674-681 ; Bibikova et al., Mol. Cell. Biol., 2001, 21, 289-297; Bibikova et al., Genetics, 2002, 161, 1169-1 175; Bibikova et al., Science, 2003, 300, 764; Porteus, M.H. and D. Baltimore, Science, 2003, 300, 763-; Alwin et al., Mol. Ther., 2005, 12, 610-617; Urnov et al., Nature, 2005, 435, 646-651 ; Porteus, M.H., Mol. Ther., 2006, 13, 438-446). Such ZFP nucleases have been used for the engineering of the IL2RG gene in human lymphoid cells (Urnov et al., Nature, 2005, 435, 646-651).

The binding specificity of Cys2-His2 type Zinc-Finger Proteins, is easy to manipulate because specificity is driven by essentially four residues per zinc finger (Pabo et al., Annu. Rev. Biochem., 2001, 70, 313-340; Jamieson et al., Nat. Rev. Drug Discov., 2003, 2, 361-368). Studies from the Pabo laboratories have resulted in a large repertoire of novel artificial ZFPs, able to bind most G/ANNG/ANNG/ANN sequences (Rebar, E.J. and CO. Pabo, Science, 1994, 263, 671-673; Kim, J.S. and CO. Pabo, Proc. Natl. Acad. Sci. U S A, 1998, 95, 2812-2817), Klug (Choo, Y. and A. Klug, Proc. Natl. Acad. Sci. USA, 1994, 91, 11163-l 1167; Isalan M. and A. Klug, Nat. Biotechnol., 2001, 19, 656-660) and Barbas (Choo, Y. and A. Klug, Proc. Natl. Acad. Sci. USA, 1994, 91, 11163-1 1167; Isalan M. and A. Klug, Nat. Biotechnol., 2001, 19, 656-660).

Nevertheless, ZFPs have serious limitations, especially for applications requiring a very high level of specificity, such as therapeutic applications. It was shown that Fokl nuclease activity in ZFP fusion proteins can act with either one recognition site or with two sites separated by variable distances via a DNA loop (Catto et al., Nucleic Acids Res., 2006, 34, 171 1-1720). Thus, the specificities of these ZFP nucleases are degenerate, as illustrated by high levels of toxicity in mammalian cells and Drosophila (Bibikova et al., Genetics, 2002, 161, 1169-1 175; Bibikova et al., Science, 2003, 300, 764-; Hockemeyer et al., Nat Biotechnol. 2009 Sep;27(9): 851-7).

Recently, a combinatorial approach was used to assemble different sets of mutations in the same new protein and to, as a result, engineer entirely redesigned meganucleases, able to cleave chosen sequences (Smith, Grizot et al. 2006; Arnould, Perez et al. 2007; Redondo, Prieto et al. 2008).

In the wild, meganucleases are essentially represented by Homing Endonucleases (HEs), a widespread family of natural meganucleases including hundreds of proteins families (Chevalier, B.S. and B.L. Stoddard, Nucleic Acids Res., 2001, 29, 3757-3774). These proteins are encoded by mobile genetic elements which propagate by a process called "homing": the endonuclease cleaves a cognate allele from which the mobile element is absent, thereby stimulating a homologous recombination event that duplicates the mobile DNA into the recipient locus. Given their exceptional cleavage properties in terms of efficacy and specificity, they could represent ideal scaffold to derive novel, highly specific endonucleases.

Although the LAGLIDADG peptide is the only conserved region among members of the largest and best characterized LAGLIDADG family, these proteins share a very similar architecture. The catalytic core is flanked by two DNA-binding domains with a perfect two¬ fold symmetry for homodimers such as I-Crel (Chevalier, et al., Nat. Struct. Biol., 2001, 8, 312-316) and I-Msol (Chevalier et al., J. Mol. Biol., 2003, 329, 253-269) and with a pseudo symmetry for monomers such as I-Scel (Moure et al., J. Mol. Biol., 2003, 334, 685-69, I-Dmol (Silva et al., J. Mol. Biol., 1999, 286, l 123-1 136) or I-Anil (Bolduc et al., Genes Dev., 2003, 17, 2875-2888).

Both monomers or both domains of monomeric proteins contribute to the catalytic core, organized around divalent cations. Just above the catalytic core, the two LAGLIDADG peptides play also an essential role in the dimerization interface. DNA binding depends on

two typical saddle-shaped αββαββα folds, sitting on the DNA major groove. Other domains can be found, for example in inteins such as Pl-Pful (Ichiyanagi et al., J. Mol. Biol., 2000, 300, 889-901) and PI-SceI (Moure et al., Nat. Struct. Biol, 2002, 9, 764-770), which protein splicing domain is also involved in DNA binding.

The making of functional chimeric meganucleases by fusing the N-terminal I-Dmol domain with an I-Crel monomer have demonstrasted the plasticity of meganucleases (Chevalier et al., Mol. Cell., 2002, 10, 895-905; Epinat et al., Nucleic Acids Res, 2003, 31, 2952-62; International PCT Applications WO 03/078619 and WO 2004/03 1346).

Different groups have used a semi-rational approach to locally alter the specificity of I-Crel (Seligman et al., Genetics, 1997, 147, 1653-1664; Sussman et al., J. Mol. Biol., 2004, 342, 31-41; International PCT Applications WO 2006/097784 and WO 2006/097853; Arnould et al., J. Mol. Biol., 2006, 355, 443-458; Rosen et al., Nucleic Acids Res., 2006, 34, 4791-4800 ; Smith et al., Nucleic Acids Res., 2006, 34, el49), I-Scel (Doyon et al., J. Am. Chem. Soc, 2006, 128, 2477-2484), PI-SceI (Gimble et al., J. Mol. Biol., 2003, 334, 993-1008 ) and I-Msol (Ashworth et al., Nature, 2006, 441, 656-659).

In addition, hundreds of I-Crel derivatives with locally altered specificity were engineered by combining the semi-rational approach and High Throughput Screening:

- Residues Q44, R68 and R70 or Q44, R68, D75 and 177 of I-Crel were mutagenized and a collection of variants with altered specificity at positions ± 3 to 5 of the DNA target (5NNN DNA target) were identified by screening (International PCT Applications WO 2006/097784 and WO 2006/097853; Arnould et al., J. Mol. Biol., 2006, 355, 443-458; Smith et al., Nucleic Acids Res., 2006, 34, e149).

- Residues K28, N30 and Q38 or N30, Y33, and Q38 or K28, Y33, Q38 and S40 of I-Crel were mutagenized and a collection of variants with altered specificity at positions ± 8 to 10 of the DNA target (10NNN DNA target) were identified by screening (Smith et al., Nucleic Acids Res., 2006, 34, e149; International PCT Applications WO 2007/060495 and WO 2007/049156).

Two different variants were combined and assembled in a functional heterodimeric endonuclease able to cleave a chimeric target resulting from the fusion of a different half of each variant DNA target sequence (Arnould et al., precited; International PCT Applications

WO 2006/097854 and WO 2007/034262). Interestingly, the novel proteins had kept proper folding and stability, high activity, and a narrow specificity.

Furthermore, residues 28 to 40 and 44 to 77 of I-CreI were shown to form two separable functional subdomains, able to bind distinct parts of a homing endonuclease half-site (Smith et al. Nucleic Acids Res., 2006, 34, e149; International PCT Applications WO 2007/049095 and WO 2007/057781).

The combination of mutations from the two subdomains of I-CreI within the same monomer allowed the design of novel chimeric molecules able to cleave a palindromic combined DNA target sequence comprising the nucleotides at positions ± 3 to 5 and ± 8 to 10 which are bound by each subdomain (Smith et al., Nucleic Acids Res., 2006, 34, e149; International PCT Applications WO 2007/060495 and WO 2007/049156), as illustrated on figure 2b.

The combination of the two former steps allows a larger combinatorial approach, involving four different subdomains. The different subdomains can be modified separately and combined to obtain an entirely redesigned meganuclease variant (heterodimer or single-chain molecule) with chosen specificity. In a first step, couples of novel meganucleases are combined in new molecules ("half-meganucleases") cleaving palindromic targets derived from the target one wants to cleave. Then, the combination of such "half-meganuclease" can result in a heterodimeric species cleaving the target of interest. The assembly of four sets of mutations into heterodimeric engineered endonucleases cleaving a model target sequence or a sequence from different genes has been described in the following patent applications for XPC gene (WO2007093918), RAG gene (WO20080 10093), HPRT gene (WO2008059382), beta-2 microglobulin gene (WO2008 102274), Rosa26 gene (WO2008 152523), Human hemoglobin beta gene (WO2009013622) and Human Interleukin-2 receptor gamma chain (WO20090 19614).

These variants can be used to cleave genuine chromosomal sequences and have paved the way for novel perspectives in several fields including gene therapy.

However, even though the base-pairs ± 1 and ± 2 do not display any contact with the protein, it has been shown that these positions are not devoid of content information (Chevalier et al., J. Mol. Biol., 2003, 329, 253-269), especially for the base-pair ± 1 and could

be a source of additional substrate specificity (Argast et al., J. Mol. Biol., 1998, 280, 345-353; Jurica et al., Mol. Cell., 1998, 2, 469-476; Chevalier, B.S. and B.L. Stoddard, Nucleic Acids Res., 2001, 29, 3757-3774). In vitro selection of cleavable I-CreI target (Argast et al., precited) randomly mutagenized, revealed the importance of these four base-pairs on protein binding and cleavage activity. It has been suggested that the network of ordered water molecules found in the active site was important for positioning the DNA target (Chevalier et al., Biochemistry, 2004, 43, 14015-14026). In addition, the extensive conformational changes that appear in this region upon I-CreI binding suggest that the four central nucleotides could contribute to the substrate specificity, possibly by sequence dependent conformational preferences (Chevalier et al., 2003, precited).

The inventors have now identified and developed novel endonucleases, such as meganucleases, targeting sequences within the WAS gene, and for instance WAS4 target sites, a site within intron 2 of the WAS gene, WAS5 a site within exon 4 of the WAS gene and WAS6, a site within intron 3 of the WAS gene as non limiting examples. The novel endonucleases and particularly the meganucleases of the invention introduce double stranded breaks within the WAS gene offering the opportunities to be used in a gene correction approach and/or in an exon knock-in approach but also in order to modify, modulate, and control WAS gene expression, to detect WAS gene expression, or to introduce transgenes into the WAS gene locus.

## BRIEF SUMMARY OF THE INVENTION

Endonucleases recognizing polynucleotide target sequences in WAS gene exons and introns have been identified and developed. These endonucleases permit the manipulation and engineering of the WAS gene locus using several strategies, including gene insertion, gene correction of defective WAS gene sequences, such as those associated with immunodeficiency or other symptoms of Wiscott-Aldrich Syndrome, or exon knock-in strategies.

The endonucleases of the invention may be employed whenever a target site recognized by a WAS meganuclease is present in a genome either naturally (site present in homolog genes in other species or site present in other locations in a genome) or artificially

introduced (i.e. use of a landing pad containing the target sequence). These endonucleases may be used in conjunction with any gene containing a recognized target sequence.

WAS gene mutations are diverse and widespread on the gene, making precise gene correction of every single mutation difficult by the identification and development of several meganucleases targeting theses different regions. Therefore, the use of an Exon Knock In (Exon KI) strategy in contrast might be a favored approach, especially for those meganucleases whose targeting sites are at the beginning of a gene such as WAS. Indeed exon KI allows the correction of mutations in the vicinity of the site as well as a functional compensation of mutations located downstream (introduction of a functional construct, shunting the mutated sequences).

In particular, the invention involves meganuclease variants that target and cleave WAS gene sequences, vectors encoding these variants, cells transformed with vectors encoding these meganuclease variants and methods for making a meganuclease variant through by expressing a polynucleotide encoding it and methods for designing meganuclease variants recognizing the WAS gene, including meganuclease variants recognizing the WAS4, WAS5 and WAS6 target DNA sequences. These variant meganucleases may also be used for therapeutic purposes and/or to investigate the function of the WAS gene, to follow its expression by introducing knock out mutations into the WAS gene or by introducing reporter genes or other genes of interest at the WAS locus. The meganuclease variants of the invention may also be used to modulate WAS expression in a cell by interaction of this gene sequence with a meganuclease, for example, to control its phenotype; to knock down or control expression of WAS in a cell or in various other therapeutic or diagnostic applications.

A particular aspect of the invention is a meganuclease that can induce double stranded breaks in any gene involved in a WAS gene associated process and particularly in the WAS gene.

Another aspect of the invention involves using such meganuclease recognizing WAS sequences to knock out or modulate WAS gene expression, particularly in appropriate cellular models ***in vitro.*** Fig. 1 illustrates WAS gene insertion, gene correction and exon knock-in strategies.

Another aspect of the invention is the use of a meganuclease recognizing WAS to introduce a gene of interest into the WAS gene or locus. The gene of interest may be a

reporter gene that permits the expression of WAS to be determined or followed over time, said reporter gene being associated or not to a nucleotidic sequence which is introduced into the genome in order to add new potentialities or properties to targeted cells, particularly in cellular models *in vitro.* Methods for determining the effects of non-WAS genes or drug compounds on WAS expression or activity may be evaluated using assays employing a reporter gene. Alternatively, the gene of interest may be a therapeutic transgene other than WAS which uses the WAS locus as a safe harbor.

Thus, a third associated aspect of the invention relates to the use of the WAS gene locus as a "landing pad" to insert or modulate the expression of genes of interest.

In addition to the preceding features, the invention further comprises other features which will emerge from the description which follows, which refers to examples illustrating the I-*CVe*I meganuclease variants and their uses according to the invention, as well as to the appended drawings. A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following Figures in conjunction with the detailed description below.

## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates various targeted approaches and strategies to precisely engineer a chosen locus. In each case, stimulation of the desired recombination event can be obtained with a meganuclease. The position of meganuclease-induced cleavage is represented by a thunderbolt.

Fig. 1A. Gene insertion strategies consist in targeting a DNA sequence to a chosen chromosomal site. For gene therapy purpose, gene insertion in a "safe harbor", i.e. a locus wherein insertion will not interfere with the expression pattern of surrounding genes should alleviate the risks of insertional mutagenesis. This can be done by homologous gene targeting (classical or stimulated by sequence specific endonucleases) or using integrases/recombinases or transposases recognizing a specific site.

Fig. IB. Gene correction is potentially the less invasive method to correct a mutation: the true reversion of the mutated gene should not be associated with ectopic genome rearrangements. Homologous gene targeting (classical, using improved donor matrices, or stimulated by sequence specific endonucleases) is the only way today to achieve this goal.

Fig. 1C. Exon knock-in is an hybrid approach wherein gene insertion is targeted to the mutated gene itself, which is then the "safe harbor". Transcription results in a mRNA that contains all the coding exons. However, this approach remains potentially more disruptive than the true gene correction strategy. The various forms of gene targeting are today the best way to achieve this goal.

Figures 2 a and b illustrate the combinatorial approach, described in International PCT applications WO 2006/097784 and WO 2006/097853 and also in Arnould, et al. (J. Mol. Biol., 2006, 355, 443-458) and Smith et al. (Nucleic Acids Res., 2006). This approach was used to entirely redesign the DNA binding domain of the I-*Cre*l protein and thereby engineer novel meganucleases with fully engineered specificity.

Fig. 3 shows WAS4 and WAS4 derived targets. The WAS4.1 target sequence (SEQ ID NO: 8) and its derivatives 10CAA_P (SEQ ID NO: 4), 5CTC_P (SEQ ID NO: 5), lOACA_P (SEQ ID NO: 6) and 5TTT_P (SEQ ID NO: 7), P stands for Palindromic) are derivatives of CI221 (SEQ ID NO: 2), found to be cleaved by previously obtained I-*Cre*I mutants.

CI221, 10CAA_P, 5CTC_P, lOACA_P and 5TTT_P were first described as 24 bp sequences, but structural data suggest that only the 22 bp are relevant for protein/DNA interaction.

WAS4.l (SEQ ID NO: 8) is the DNA sequence located in intron 2 of the WAS gene at positions 1518 to 1541 of entry NC000023. WAS4.2 (SEQ ID NO: 9) differs from WAS4.1 at positions -2;-l;+l ;+2, where *l-CreI* cleavage site (GTAC) substitutes the corresponding WAS4.1 sequence. WAS4.3 (SEQ ID NO: 10) is the palindromic sequence derived from the left part of WAS4.2, and WAS4.4 (SEQ ID NO: 11) is the palindromic sequence derived from the right part of WAS4.2. WAS4.5 (SEQ ID NO: 12) is the palindromic sequence derived from the left part of WAS4.1, and WAS4.6 (SEQ ID NO: 13) is the palindromic sequence derived from the right part of WAS4.1.

Fig. 4 compares the activities of WAS4 variants in CHO SSA assay as described in Example 1 below. Activity cleavage in CHO cells of single chain heterodimer pCLS3662, pCLS3702 and pCLS3703 compared to IScel (pCLS1090) and SCOH-RAG-CLS (pCLS2222) meganucleases as positive controls. The empty vector control (pCLS1069) has also been tested on each target. Plasmid pCLS1728 contains control RAG1.10.1 target sequence. In figure 4, the correspondence of the line graphs at their right ends to the legend (graph:legend) on the right is as follows: graph 1 (top) : 3; 2 : 5; 3 : 6; 4 : 4; 5 : 2 and 6 : 1.

Fig. 5 shows WAS5 and WAS5 derived targets. The WAS5.1 target sequence (SEQ ID NO: 18) and its derivatives 10GGG_P (SEQ ID NO: 14), 5CTC_P (SEQ ID NO: 15), 10CTT_P (SEQ ID NO: 16) and 5CCT_P (SEQ ID NO: 17), P stands for Palindromic) are derivatives of C1221 (SEQ ID NO: 2), found to be cleaved by previously obtained \-*Cre*\ mutants.

C1221, 10GGG_P, 5CTC_P, 10CTT_P and 5CCT_P were first described as 24 bp sequences, but structural data suggest that only the 22 bp are relevant for protein/DNA interaction.

WAS5.1 (SEQ ID NO: 18) is the DNA sequence located in exon 4 of the WAS gene at positions 1965 to 1988 of entry NC000023. WAS5.2 (SEQ ID NO: 19) differs from WAS5.1 at positions -2;-l ;+l;+2, where I-*Cre*I cleavage site (GTAC) substitutes the corresponding WAS5.1 sequence. WAS5.3 (SEQ ID NO: 20) is the palindromic sequence derived from the left part of WAS5.2, and WAS5.4 (SEQ ID NO: 21) is the palindromic sequence derived from the right part of WAS5.2. WAS5.5 (SEQ ID NO: 22) is the palindromic sequence derived from the left part of WAS5.1, and WAS5.6 (SEQ ID NO: 23) is the palindromic sequence derived from the right part of WAS5.1.

Fig. 6 compares the activities of WAS5 variants in the CHO SSA assay as described in Example 2 below. Activity cleavage in CHO cells of single chain heterodimer pCLS3751, pCLS3752, pCLS3753, pCLS3754, pCLS3755, pCLS3756, pCLS3757, pCLS3758 and pCLS3759 compared to IScel (pCLS1090) and SCOH-RAG-CLS (pCLS2222) meganucleases as positive controls. The empty vector control (pCLS1069) has also been tested on each target. Plasmid pCLS1728 contains control RAG1.10.1 target sequence. In figure 6, the correspondence of the line graphs at their right ends to the legend (graph:legend)

on the right is as follows: graph 1 (top) : 12; 2: 2; 3: 3; 4: 9; 5: 10; 6: 4; 7: 11; 8: 8 and 5; 10: 6; 11: 1 and 12: 7.

Fig. 7 shows WAS6 and WAS6 derived targets. The WAS6. 1 target sequence (SEQ ID NO: 28) and its derivatives 10GGC_P (SEQ ID NO: 24), 5CAC_P (SEQ ID NO: 25), 10TCC_P (SEQ ID NO: 26) and 5ACC_P (SEQ ID NO: 27), P stands for Palindromic) are derivatives of C1221 (SEQ ID NO: 2), found to be cleaved by previously obtained I-*Cre*I mutants.

C1221, 10GGC_P, 5CAC_P, 10TCC_P and 5ACC_P were first described as 24 bp sequences, but structural data suggest that only the 22 bp are relevant for protein/DNA interaction.

WAS6.1 (SEQ ID NO: 28) is the DNA sequence located in intron 3 of the WAS gene at positions 1835 to 1858 of entry NC000023. WAS6.2 (SEQ ID NO: 29) differs from WAS6.1 at positions -2;-l;+l;+2, where I-*Cre*l cleavage site (GTAC) substitutes the corresponding WAS6.1 sequence. WAS6.3 (SEQ ID NO: 30) is the palindromic sequence derived from the left part of WAS6.2, and WAS6.4 (SEQ ID NO: 31) is the palindromic sequence derived from the right part of WAS6.2. WAS6.5 (SEQ ID NO: 32) is the palindromic sequence derived from the left part of WAS6.1, and WAS6.6 (SEQ ID NO: 33) is the palindromic sequence derived from the right part of WAS6. 1.

Fig. 8 shows a vector map of pCLS1072.

Fig. 9 shows a vector map of pCLS1090.

Fig. 10 shows a vector map of pCLS2222.

Fig. 11 shows a vector map of pCLSl 853.

Fig. 12 shows a vector map of pCLS1107.

Fig. 13 shows a vector map of pCLS0002.

Fig. 14 shows a vector map of pCLS1069.

Fig. 15 shows a vector map of pCLS1058.

Fig. 16 shows a vector map of pCLS1728.

Fig. 17 shows a vector map of pCLS5057

Fig. 18 shows *In vivo* characterizations of both pCLS3756 and pCLS9067 WAS5 meganucleases (respectively, SEQ ID NO: 85 and 94). **A.** CHO-K1 cells were co-transfected with the reporter plasmid measuring SSA activity containing the WAS5 target and increasing amounts of pCLS3756 and pCLS9067. Activity of I-SceI is shown as a control. **B.** CHO-K1 cells were co-transfected with increasing amounts of meganuclease expression vector and a constant amount of a plasmid encoding for GFP. Cell survival is expressed as a percentage of cells still expressing GFP 6 days post transfection. MegaX, a meganuclease with a relaxed specificity, is shown as a positive control for toxicity. **C.** Gene targeting experiment: 293 cells were co-transfected with the repair plasmid pCLS5057 ( SEQ ID NO: 78) and pCLS3756, pCLS9067 or an empty vector (mock).

## DETAILED DESCRIPTION OF THE INVENTION

The present invention concerns new classes of endonucleases recognizing polynucleotide target sequences in WAS gene. These endonucleases permit the manipulation and engineering of the WAS gene locus using several strategies, including gene insertion, gene correction of defective WAS gene sequences, such as those associated with immunodeficiency or other symptoms of Wiscott-Aldrich Syndrome, or exon knock-in strategies.

The endonucleases of the invention may be employed whenever a target site recognized by a WAS meganuclease is present in a genome either naturally (site present in homolog genes in other species or site present in other locations in a genome) or artificially introduced (i.e. use of a landing pad containing the target sequence). These endonucleases may be used in conjunction with any gene containing a recognized target sequence.

WAS gene mutations are diverse and widespread on the gene, making precise gene correction of every single mutation difficult by the identification and development of several meganucleases targeting theses different regions difficult. Therefore, the use of an Exon Knock In (Exon KI) strategy in contrast might be a favored approach, especially for those meganucleases whose targeting sites are at the beginning of a gene such as WAS. Indeed exon

KI allows the correction of mutations in the vicinity of the site as well as a functional compensation of mutations located downstream (introduction of a functional construct, shunting the mutated sequences). Since the WAS locus could be used for the expression of reporter and genes of interest such as in appropriate culture cell models, some meganuclease targeting sequences in exons (such as WAS5 as a non limiting example) or in introns (such as WAS4 and WAS6 as non limiting examples) are useful for the integration of knock in matrix by homologous recombination. Such a KI matrix can be built using sequences homologous to the targeted locus added of the gene of interest with or without regulation elements.

In particular, the invention involves meganuclease variants that target and cleave WAS gene sequences, vectors encoding these variants, cells transformed with vectors encoding these meganuclease variants and methods for making a meganuclease variant through by expressing a polynucleotide encoding it and methods for designing meganuclease variants recognizing the WAS gene, including meganuclease variants recognizing the WAS4, WAS5 and WAS6 target sequences. These variant meganucleases may be used for therapeutic purposes and/or to investigate the function of the WAS gene, to follow its expression by introducing knock out mutations into the WAS gene or by introducing reporter genes or other genes of interest at the WAS locus. The meganuclease variants of the invention may also be used to modulate WAS expression in a cell by interaction of this gene sequence with a meganuclease, for example, to control its phenotype; to knock down or control expression of WAS in a cell such as a tumor cell, or in various other therapeutic or diagnostic applications.

A particular aspect of the invention is a meganuclease that can induce double stranded breaks in any gene involved in a WAS gene associated process and particularly in the WAS gene.

Another aspect of the invention involves using such a meganuclease recognizing WAS sequences to correct WAS pathological mutations. The coding sequence can be corrected upon meganuclease cleavage and recombination with a repair matrix (Fig. IB).

Another aspect of the invention involves the use of meganuclease recognizing WAS gene to knock out its expression, particularly in appropriate cellular models *in vitro.* Different strategies can be implemented for inactivating WAS gene. It can be inactivated by Non Homologous End Joining (NHEJ) using a meganuclease targeting a sequence without a repair matrix (non shown). Meganuclease targeting the WAS5 sequence is such an enzyme. In that case, no matrix is needed. Some exons can be deleted by the action of one meganuclease

supplied by a Knocking out DNA matrix. Meganucleases recognizing WAS4, WAS5 or WAS6 sequences are useful. A second sub-type of knock-out strategy consists in the replacement of a large region within WAS gene by the action of two meganucleases (example: WAS4 + WAS6) and a KO matrix can be used for the deletion of large sequences (not shown). Such a KO matrix can be built using sequences deleted of the targeted exon as well as some mutated exons.

Another aspect of the invention can be the use of a meganuclease recognizing WAS to introduce a gene of interest into the WAS gene or locus. The gene of interest may be a reporter gene that permits the expression of WAS to be determined or followed over time, said reporter gene being associated or not to a nucleotidic sequence which is introduced into the genome in order to add new potentialities or properties to targeted cells, particularly in cellular models *in vitro.* Methods for determining the effects of genes other than WAS or drug compounds on WAS expression or activity may be evaluated using assays employing a reporter gene. Alternatively, the gene of interest may be a therapeutic transgene other than WAS which uses the WAS locus as a safe harbor.

Thus, a third associated aspect of the invention relates to the use of the WAS gene as a "landing pad" to insert or modulate the expression of genes of interest.

In addition to the preceding features, the invention further comprises other features which will emerge from the description which follows, which refers to examples illustrating the I-Crel meganuclease variants and their uses according to the invention, as well as to the appended drawings. A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by reference to the following Figures in conjunction with the detailed description below.

Meganucleases directed against WAS gene will therefore represent a tool of choice for therapeutic purposes and several other applications which will permit to better understand the effects of WAS gene mutations and WAS gene function and thus overcome actual problems in WAS cell therapy and drug screening studies.

As mentioned above certain aspects of the invention reflect different strategies for modulating, modifying or controlling WAS gene expression that can be implemented with the WAS gene recognizing meganucleases of the invention.

I-CreI variants of the present invention were created using the combinatorial approach illustrated in Figure 2b and described in International PCT applications WO 2006/097784 and WO 2006/097853, and also in Arnould et al. (J. Mol. Biol., 2006, 355, 443-458) and Smith et al. (Nucleic Acids Res., 2006), allowing to redesign the DNA binding domain of the I-CreI protein and thereby engineer novel meganucleases with fully engineered specificity.

The cleavage activity of the variant according to the invention may be performed by any well-known, in vitro or in vivo cleavage assay, such as those described in the International PCT Application WO 2004/067736; Epinat et al., Nucleic Acids Res., 2003, 31, 2952-2962; Chames et al., Nucleic Acids Res., 2005, 33, e178; Arnould et al., J. Mol. Biol., 2006, 355, 443-458, and Arnould et al., J. Mol. Biol., 2007, 371, 49-65. For example, the cleavage activity of the variant of the invention may be measured by a direct repeat recombination assay, in yeast or mammalian cells, using a reporter vector. The reporter vector comprises two truncated, non-functional copies of a reporter gene (direct repeats) and the genomic (non-palindromic) DNA target sequence within the intervening sequence, cloned in yeast or in a mammalian expression vector. Usually, the genomic DNA target sequence comprises one different half of each (palindromic or pseudo-palindromic) parent homodimeric I-CreI meganuclease target sequence. Expression of the heterodimeric variant results in a functional endonuclease which is able to cleave the genomic DNA target sequence. This cleavage induces homologous recombination between the direct repeats, resulting in a functional reporter gene, whose expression can be monitored by an appropriate assay. The cleavage activity of the variant against the genomic DNA target may be compared to wild type I-CreI or I-SceI activity against their natural target.

Possibly or not, at least two rounds of selection/screening are performed according to the process illustrated Arnould et al., J. Mol. Biol., 2007, 371, 49-65. In the first round, one of the monomers of the heterodimer is mutagenised, co-expressed with the other monomer to form heterodimers, and the improved monomers $Y^+$ are selected against the target from the gene of interest. In the second round, the other monomer (monomer X) is mutagenised, co-expressed with the improved monomers $Y^+$ to form heterodimers, and selected against the target from the gene of interest to obtain meganucleases ($X^+$ $Y^+$) with improved activity. The mutagenesis may be random-mutagenesis or site-directed mutagenesis on a monomer or on a pool of monomers, as indicated above. Both types of mutagenesis are advantageously

combined. Additional rounds of selection/screening on one or both monomers may be performed to improve the cleavage activity of the variant.

In a preferred embodiment of said variant, said substitution(s) in the subdomain situated from positions 44 to 77 of I-CreI are at positions 44, 68, 70, 75 and/or 77.

In another preferred embodiment of said variant, said substitution(s) in the subdomain situated from positions 28 to 40 of I-CreI are at positions 28, 30, 32, 33, 38 and/or 40.

In another preferred embodiment of said variant, it comprises one or more mutations in I-CreI monomer(s) at positions of other amino acid residues that contact the DNA target sequence or interact with the DNA backbone or with the nucleotide bases, directly or via a water molecule; these residues are well-known in the art (Jurica et al., Molecular Cell., 1998, 2, 469-476; Chevalier et al., J. Mol. Biol., 2003, 329, 253-269). In particular, additional substitutions may be introduced at positions contacting the phosphate backbone, for example in the final C-terminal loop (positions 137 to 143; Prieto et al., Nucleic Acids Res., Epub 22 April 2007).

Preferably said residues are involved in binding and cleavage of said DNA cleavage site. More preferably, said residues are at positions 138, 139, 142 or 143 of I-CreI. Two residues may be mutated in one variant provided that each mutation is in a different pair of residues chosen from the pair of residues at positions 138 and 139 and the pair of residues at positions 142 and 143. The mutations which are introduced modify the interaction(s) of said amino acid(s) of the final C-terminal loop with the phosphate backbone of the I-CreI site. Preferably, the residue at position 138 or 139 is substituted by a hydrophobic amino acid to avoid the formation of hydrogen bonds with the phosphate backbone of the DNA cleavage site. For example, the residue at position 138 is substituted by an alanine or the residue at position 139 is substituted by a methionine. The residue at position 142 or 143 is advantageously substituted by a small amino acid, for example a glycine, to decrease the size of the side chains of these amino acid residues.

More preferably, said substitution in the final C-terminal loop modify the specificity of the variant towards the nucleotide at positions ± 1 to 2, ± 6 to 7 and/or ± 11 to 12 of the I-CreI site.

In another preferred embodiment of said variant, it comprises one or more additional mutations that improve the binding and/or the cleavage properties of the variant towards the DNA target sequence from the WAS gene. The additional residues which are mutated may be on the entire I-CreI sequence, and in particular in the C-terminal half of I-CreI (positions 80 to 163). Both 1-CreI monomers are advantageously mutated; the mutation(s) in each monomer may be identical or different. For example, the variant comprises one or more additional substitutions at positions: 2, 7, 8, 19, 43, 54, 61, 80, 81, 96, 105 and 132. Said substitutions are advantageously selected from the group consisting of: N2S, K7E, E8K, G19S, F43L, F54L, E61R, E80K, 18IT, K96E, V105A and I132V. More preferably, the variant comprises at least one substitution selected from the group consisting of: N2S, K7E, E8K, G19S, F43L, F54L, E61R, E80K, 18IT, K96E, V105A and I132V. The variant may also comprise additional residues at the C-terminus. For example a glycine (G) and/or a proline (P) residue may be inserted at positions 164 and 165 of I-CreI, respectively.

According to a preferred embodiment, said additional mutation in said variant further impairs the formation of a functional homodimer. More preferably, said mutation is the G19S mutation. The G19S mutation is advantageously introduced in one of the two monomers of a heterodimeric I-CreI variant, so as to obtain a meganuclease having enhanced cleavage activity and enhanced cleavage specificity. In addition, to enhance the cleavage specificity further, the other monomer may carry a distinct mutation that impairs the formation of a functional homodimer or favors the formation of the heterodimer.

In another preferred embodiment of said variant, said substitutions are replacement of the initial amino acids with amino acids selected from the group consisting of: A, D, E, G, H, K, N, P, Q, R, S, T, Y, C, V, L, M, F, I and W.

The variant of the invention may be derived from the wild-type I-CreI (SEQ ID NO: 1) or an I-CreI scaffold protein having at least 85 % identity, preferably at least 90 % identity, more preferably at least 95 % identity with SEQ ID NO:l, such as the scaffold called I-CreI N75 ( amino acids; SEQ ID NO: 3) having the insertion of an alanine at position 2, and the insertion of AAD at the C-terminus (positions 164 to 166) of the I-CreI sequence. In the present Patent Application all the I-CreI variants described comprise an additional Alanine after the first Methionine of the wild type I-CreI sequence (SEQ ID NO: 1). These variants also comprise two additional Alanine residues and an Aspartic Acid residue after the final Proline of the wild type I-CreI sequence. These additional residues do not affect the properties

of the enzyme and to avoid confusion these additional residues do not affect the numeration of the residues in I-CreI or a variant referred in the present Patent Application, as these references exclusively refer to residues of the wild type I-CreI enzyme (SEQ ID NO: 1) as present in the variant, so for instance residue 2 of I-CreI is in fact residue 3 of a variant which comprises an additional Alanine after the first Methionine.

In addition, the variants of the invention may include one or more residues inserted at the $NH_2$ terminus and/or COOH terminus of the sequence. For example, a tag (epitope or polyhistidine sequence) is introduced at the $NH_2$ terminus and/or COOH terminus; said tag is useful for the detection and/or the purification of said variant. The variant may also comprise a nuclear localization signal (NLS); said NLS is useful for the importation of said variant into the cell nucleus. The NLS may be inserted just after the first methionine of the variant or just after an N-terminal tag.

The variant according to the present invention may be a homodimer which is able to cleave a palindromic or pseudo-palindromic DNA target sequence.

Alternatively, said variant is a heterodimer, resulting from the association of a first and a second monomer having different substitutions at positions 28 to 40 and 44 to 77 of I-CreI, said heterodimer being able to cleave a non-palindromic DNA target sequence from the WAS gene.

In particular said heterodimer variant is composed by one of the possible associations between variants constituting N-terminal and C-terminal monomers of single chain molecules from the group consisting of SEQ ID NO: 35 to SEQ ID NO: 37, SEQ ID NO: 38 to SEQ ID NO: 46 and SEQ ID NO: 47 to SEQ ID NO: 50.

As shown in tables I, II and III and examples, sequences of the plasmids encoding the above-mentioned single chain molecules of SEQ ID NO: 35 to SEQ ID NO: 37, SEQ ID NO: 38 to SEQ ID NO: 46 and SEQ ID NO: 47 to SEQ ID NO: 50 are given , respectively, in SEQ ID NO: 51 to SEQ ID NO: 53, SEQ ID NO: 54 to SEQ ID NO: 62 and SEQ ID NO: 63 to SEQ ID NO: 66.

The DNA target sequences are situated in the WAS Open Reading Frame (ORF) and these sequences cover all the WAS ORF. In particular, said DNA target sequences recognized by the variant of the present invention and derivatives are selected from the group consisting of the SEQ ID NO: 4 to SEQ ID NO: 33, as shown in figures 3, 5 and 7.

The sequence of each I-Crel variant is defined by the mutated residues at the indicated positions. The positions are indicated by reference to I-Crel sequence (SEQ ID NO: 1) ; I-Crel has N, S, Y, Q, S, Q, R, R, D, I and E at positions 30, 32, 33, 38, 40, 44, 68, 70, 75, 77 and 80 respectively.

Each monomer (first monomer and second monomer) of the heterodimeric variant according to the present invention may also be named with a letter code, after the eleven residues at positions 28, 30, 32, 33, 38, 40, 44, 68 and 70, 75 and 77 and the additional residues which are mutated, as indicated above. For example, the mutations 7E17A33C40Q44K68A96E132N in the N-terminal monomer constituting a single chain molecule of the present invention targeting the WAS2 target (SEQ ID NO: 35).

In the present invention, for a given DNA target, ".2" derivative target sequence differs from the initial genomic target at positions -2, -1, +1, +2, where I-Crel cleavage site (GTAC) substitutes the corresponding sequence at these positions of said initial genomic target. ".3" derivative target sequence is the palindromic sequence derived from the left part of said ".2" derivative target sequence. ".4" derivative target sequence is the palindromic sequence derived from the right part of said ".2" derivative target sequence. ".5" derivative target sequence is the palindromic sequence derived from the left part of the initial genomic target. ".6" derivative is the palindromic sequence derived from the left part of the initial genomic target.

In the present invention, a "N-terminal monomer" constituting one of the monomers of a single chain molecule, refers to a variant able to cleave ".3" or ".5" palindromic sequence. In the present invention, a "C-terminal monomer" constituting one of the monomers of a single chain molecule, refers to a variant able to cleave ".4" or ".6" palindromic sequence.

The heterodimeric variant as defined above may have only the amino acid substitutions as indicated above. In this case, the positions which are not indicated are not mutated and thus correspond to the wild-type I-Crel (SEQ ID NO: 1).

The invention encompasses I-Crel variants having at least 85 % identity, preferably at least 90 % identity, more preferably at least 95 % (96 %, 97 %, 98 %, 99 %) identity with the sequences as defined above, said variant being able to cleave a DNA target from the WAS gene.

The heterodimeric variant is advantageously an obligate heterodimer variant having at least one pair of mutations corresponding to residues of the first and the second monomers which make an intermolecular interaction between the two I-CreI monomers, wherein the first mutation of said pair(s) is in the first monomer and the second mutation of said pair(s) is in the second monomer and said pair(s) of mutations prevent the formation of functional homodimers from each monomer and allow the formation of a functional heterodimer, able to cleave the genomic DNA target from the WAS gene.

To form an obligate heterodimer, the monomers have advantageously at least one of the following pairs of mutations, respectively for the first monomer and the second monomer:

a) the substitution of the glutamic acid at position 8 with a basic amino acid, preferably an arginine (first monomer) and the substitution of the lysine at position 7 with an acidic amino acid, preferably a glutamic acid (second monomer); the first monomer may further comprise the substitution of at least one of the lysine residues at positions 7 and 96, by an arginine,

b) the substitution of the glutamic acid at position 61 with a basic amino acid, preferably an arginine (first monomer) and the substitution of the lysine at position 96 with an acidic amino acid, preferably a glutamic acid (second monomer); the first monomer may further comprise the substitution of at least one of the lysine residues at positions 7 and 96, by an arginine,

c) the substitution of the leucine at position 97 with an aromatic amino acid, preferably a phenylalanine (first monomer) and the substitution of the phenylalanine at position 54 with a small amino acid, preferably a glycine (second monomer); the first monomer may further comprise the substitution of the phenylalanine at position 54 by a tryptophane and the second monomer may further comprise the substitution of the leucine at position 58 or lysine at position 57, by a methionine, and

d) the substitution of the aspartic acid at position 137 with a basic amino acid, preferably an arginine (first monomer) and the substitution of the arginine at position 51 with an acidic amino acid, preferably a glutamic acid (second monomer).

For example, the first monomer may have the mutation D137R and the second monomer, the mutation R51D. The obligate heterodimer meganuclease comprises advantageously, at least two pairs of mutations as defined in a), b), c) or d), above; one of the

pairs of mutation is advantageously as defined in c) or d). Preferably, one monomer comprises the substitution of the lysine residues at positions 7 and 96 by an acidic amino acid (aspartic acid (D) or glutamic acid (E)), preferably a glutamic acid (K7E and K96E) and the other monomer comprises the substitution of the glutamic acid residues at positions 8 and 61 by a basic amino acid (arginine (R) or lysine (K); for example, E8K and E61R). More preferably, the obligate heterodimer meganuclease, comprises three pairs of mutations as defined in a), b) and c), above.

The obligate heterodimer meganuclease consists advantageously of a first monomer (A) having at least the mutations (i) E8R, E8K or E8H, E61R, E61K or E61H and L97F, L97W or L97Y; (ii) K7R, E8R, E61R, K96R and L97F, or (iii) K7R, E8R, F54W, E61R, K96R and L97F and a second monomer (B) having at least the mutations (iv) K7E or K7D, F54G or F54A and K96D or K96E; (v) K7E, F54G, L58M and K96E, or (vi) K7E, F54G, K57M and K96E. For example, the first monomer may have the mutations K7R, E8R or E8K, E61R, K96R and L97F or K7R, E8R or E8K, F54W, E61R, K96R and L97F and the second monomer, the mutations K7E, F54G, L58M and K96E or K7E, F54G, K57M and K96E. The obligate heterodimer may comprise at least one NLS and/or one tag as defined above; said NLS and/or tag may be in the first and/or the second monomer.

The subject-matter of the present invention is also a single-chain chimeric meganuclease (fusion protein) derived from an I-CreI variant as defined above. The single-chain meganuclease may comprise two I-CreI monomers, two I-CreI core domains (positions 6 to 94 of I-CreI) or a combination of both. Preferably, the two monomers/core domains or the combination of both, are connected by a peptidic linker.

It is understood that the scope of the present invention also encompasses the I-CreI variants per se, including heterodimers, obligate heterodimers, single chain meganucleases as non-limiting examples, able to cleave one of the target sequences in WAS gene.

It is also understood that the scope of the present invention also encompasses the I-CreI variants as defined above that target equivalent sequences in WAS gene of eukaryotic organisms other than human, preferably mammals, more preferably a laboratory rodent (mice, rat, guinea-pig), or a rabbit, a cow, pig, horse or goat, those sequences being identified by the man skilled in the art in public databank like NCBI.

It is also understood that the scope of the present invention also encompasses endonucleases derived from a TALE-nuclease (TALEN) as explained below. Particularly, endonucleases according to the invention can be a fusion between a DNA-binding domain derived from a Transcription Activator Like Effector (TALE) and one or two catalytic domains having endonuclease activity. Such endonucleases can target sequences within TERT gene locus according to the present invention selected from the group consisting of SEQ ID NO: 54 to 64 as non-limiting example.

The subject matter of the present invention is also a polynucleotide fragment encoding a variant or a single-chain chimeric meganuclease as defined above; said polynucleotide may encode one monomer of a homodimeric or heterodimeric variant, or two domains/monomers of a single-chain chimeric meganuclease. It is understood that the subject matter of the present invention is also a polynucleotide fragment encoding one of the variant species as defined above, obtained by any method well known in the art.

The subject matter of the present invention is also a recombinant vector for the expression of a variant or a single-chain meganuclease according to the invention. The recombinant vector comprises at least one polynucleotide fragment encoding a variant or a single-chain meganuclease, as defined above. In a preferred embodiment, said vector comprises two different polynucleotide fragments, each encoding one of the monomers of a heterodimeric variant.

A vector, which can be used in the present invention, includes, but is not limited to, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule, which may consist of chromosomal, non-chromosomal, semi-synthetic or synthetic nucleic acids. Preferred vectors are those capable of autonomous replication (episomal vector) and/or expression of nucleic acids to which they are linked (expression vectors). Large numbers of suitable vectors are known to those skilled in the art and commercially available.

Viral vectors include retrovirus, adenovirus, parvovirus (e. g. adeno-associated viruses), coronavirus, negative strand RNA viruses such as orthomyxovirus (e. g., influenza virus), rhabdovirus (e. g., rabies and vesicular stomatitis virus), paramyxovirus (e. g. measles and Sendai), positive strand RNA viruses such as picornavirus and alphavirus, and double-stranded DNA viruses including adenovirus, herpesvirus (e. g., Herpes Simplex virus types 1 and 2, Epstein-Barr virus, cytomegalovirus), and poxvirus (e. g., vaccinia, fowlpox and canarypox). Other viruses include Norwalk virus, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus, and hepatitis virus, for example. Examples of retroviruses include:

avian leukosis-sarcoma, mammalian C-type, B-type viruses, D type viruses, HTLV-BLV group, lentivirus (particularly self-inactivating lentiviral vectors), spumavirus (Coffin, J. M., Retroviridae: The viruses and their replication, In Fundamental Virology, Third Edition, B. N. Fields, et al., Eds., Lippincott-Raven Publishers, Philadelphia, 1996).

Vectors can comprise selectable markers, for example: neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, Glutamine Synthetase, and hypoxanthine-guanine phosphoribosyl transferase for eukaryotic cell culture; TRP1, URA3 and LEU2 for S. cerevisiae; tetracycline, rifampicin or ampicillin resistance in E. coli.

Preferably, said vectors are expression vectors, wherein the sequence(s) encoding the variant/single-chain meganuclease of the invention is placed under control of appropriate transcriptional and translational control elements to permit production or synthesis of said variant. Therefore, said polynucleotide is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a promoter operatively linked to said polynucleotide, a ribosome-binding site, an RNA-splicing site (when genomic DNA is used), a polyadenylation site and a transcription termination site. It also can comprise an enhancer. Selection of the promoter will depend upon the cell in which the polypeptide is expressed. Preferably, when said variant is a heterodimer, the two polynucleotides encoding each of the monomers are included in one vector, which is able to drive the expression of both polynucleotides, simultaneously. Suitable promoters include tissue specific and/or inducible promoters. Examples of inducible promoters are: eukaryotic metallothionine promoter which is induced by increased levels of heavy metals, prokaryotic lacZ promoter which is induced in response to isopropyl-j3-D-fhiogalacto-pyranoside (IPTG) and eukaryotic heat shock promoter which is induced by increased temperature. Examples of tissue specific promoters are skeletal muscle creatine kinase, prostate-specific antigen (PSA), a-antitrypsin protease, human surfactant (SP) A and B proteins, β-casein and acidic whey protein genes.

According to another advantageous embodiment of said vector, it includes a targeting construct comprising sequences sharing homologies with the region surrounding the genomic DNA cleavage site as defined above.

For instance, said sequence sharing homologies with the regions surrounding the genomic DNA cleavage site of the variant is a fragment of the WAS gene. Alternatively, the

vector coding for an I-CreI variant/single-chain meganuclease and the vector comprising the targeting construct are different vectors.

More preferably, the targeting DNA construct comprises:

a) sequences sharing homologies with the region surrounding the genomic DNA cleavage site as defined above, and
b) a sequence to be introduced flanked by sequences as in a) or included in sequences as in a).

Preferably, homologous sequences of at least 50 bp, preferably more than 100 bp and more preferably more than 200 bp are used. Therefore, the targeting DNA construct is preferably from 200 bp to 6000 bp, more preferably from 1000 bp to 2000 bp. Indeed, shared DNA homologies are located in regions flanking upstream and downstream the site of the break and the DNA sequence to be introduced should be located between the two arms. The sequence to be introduced may be any sequence used to alter the chromosomal DNA in some specific way including a sequence used to repair a mutation in the WAS gene, restore a functional WAS gene in place of a mutated one, modify a specific sequence in the WAS gene, to attenuate or activate the WAS gene, to inactivate or delete the WAS gene or part thereof, to introduce a mutation into a site of interest or to introduce an exogenous gene or part thereof. Such chromosomal DNA alterations are used for genome engineering (animal models/recombinant cell lines) or genome therapy (gene correction or recovery of a functional gene). The targeting construct comprises advantageously a positive selection marker between the two homology arms and eventually a negative selection marker upstream of the first homology arm or downstream of the second homology arm. The marker(s) allow(s) the selection of cells having inserted the sequence of interest by homologous recombination at the target site.

The sequence to be introduced is a sequence which repairs a mutation in the WAS gene (gene correction or recovery of a functional gene), for the purpose of genome therapy. For correcting the WAS gene, cleavage of the gene occurs in the vicinity of the mutation, preferably, within 500 bp of the mutation. The targeting construct comprises a WAS gene fragment which has at least 200 bp of homologous sequence flanking the target site (minimal repair matrix) for repairing the cleavage, and includes a sequence encoding a portion of wild-type WAS gene corresponding to the region of the mutation for repairing the mutation.

Consequently, the targeting construct for gene correction comprises or consists of the minimal repair matrix; it is preferably from 200 bp to 6000 bp, more preferably from 1000 bp to 2000 bp. Preferably, when the cleavage site of the variant overlaps with the mutation the repair matrix includes a modified cleavage site that is not cleaved by the variant which is used to induce said cleavage in the WAS gene and a sequence encoding wild-type WAS gene that does not change the open reading frame of the WAS gene.

Alternatively, for the generation of knock-in cells/animals, the targeting DNA construct may comprise flanking regions corresponding to WAS gene fragments which has at least 200 bp of homologous sequence flanking the target site of the I-CreI variant for repairing the cleavage, an exogenous gene of interest within an expression cassette and eventually a selection marker such as the neomycin resistance gene.

For the insertion of a sequence, DNA homologies are generally located in regions directly upstream and downstream to the site of the break (sequences immediately adjacent to the break; minimal repair matrix). However, when the insertion is associated with a deletion of ORF sequences flanking the cleavage site, shared DNA homologies are located in regions upstream and downstream the region of the deletion.

Alternatively, for restoring a functional gene cleavage of the gene occurs in the vicinity or upstream of a mutation. Preferably, said mutation is the first known mutation in the sequence of the gene, so that all the downstream mutations of the gene can be corrected simultaneously. The targeting construct comprises the exons downstream of the cleavage site fused in frame (as in the cDNA) and with a polyadenylation site to stop transcription in 3'. The sequence to be introduced (exon knock-in construct) is flanked by introns or exons sequences surrounding the cleavage site, so as to allow the transcription of the engineered gene (exon knock-in gene) into a mRNA able to code for a functional protein. For example, the exon knock-in construct is flanked by sequences upstream and downstream of the cleavage site, from a minimal repair matrix as defined above.

The subject matter of the present invention is also a targeting DNA construct as defined above.

The subject matter of the present invention is also a composition characterized in that it comprises at least one meganuclease as defined above (variant or single-chain chimeric

meganuclease) and/or at least one expression vector encoding said meganuclease, as defined above. Preferably, said composition is a pharmaceutical composition.

In a preferred embodiment of said composition, it comprises a targeting DNA construct, as defined above. Preferably, said targeting DNA construct is either included in a recombinant vector or it is included in an expression vector comprising the polynucleotide(s) encoding the meganuclease according to the invention.

The subject-matter of the present invention is further the use of a meganuclease as defined above, one or two polynucleotide(s), preferably included in expression vector(s), for repairing mutations of the WAS gene.

The subject-matter of the present invention is also further a method of treatment of a genetic disease caused by a mutation in WAS gene comprising administering to a subject in need thereof an effective amount of at least one variant encompassed in the present invention.

According to an advantageous embodiment of said use, it is for inducing a double-strand break in a site of interest of the WAS gene comprising a genomic DNA target sequence, thereby inducing a DNA recombination event, a DNA loss or cell death.

According to the invention, said double-strand break is for: repairing a specific sequence in the WAS gene, modifying a specific sequence in the WAS gene, restoring a functional WAS gene in place of a mutated one, attenuating or activating the WAS gene, introducing a mutation into a site of interest of the WAS gene, introducing an exogenous gene or a part thereof, inactivating or deleting the WAS gene or a part thereof, translocating a chromosomal arm, or leaving the DNA unrepaired and degraded.

Therefore, in another aspect of the present invention, the inventors have found that endonucleases variants targeting WAS gene can be used for inserting therapeutic transgenes other than WAS at WAS gene locus, using this locus as a safe harbor locus. In other terms, the invention relates to a mutant endonuclease capable of cleaving a target sequence in the WAS gene, for use in safely inserting a transgene, wherein said disruption or deletion of said locus does not modify expression of genes located outside of said locus.

The subject-matter of the present invention is also further a method of treatment of a genetic disease caused by a mutation in a gene other than WAS gene comprising

administering to a subject in need thereof an effective amount of at least one variant encompassed in the present invention.

The skilled in the art can easily verify whether disruption or deletion of a locus modifies expression of neighboring genes located outside of said locus using proteomic tools. Many protein expression profiling arrays suitable for such an analysis are commercially available. By "neighboring genes" is meant the 1, 2, 5, 10, 20 or 30 genes that are located at each end of the WAS gene locus.

In a derived main aspect of the present invention, the inventors have found that the WAS locus could be used as a landing pad to insert and express genes of interest (GOIs) other than therapeutics. In this aspect, inventors have found that genetic constructs containing a GOI could be integrated into the genome at the WAS gene via meganuclease-induced recombination by specific meganuclease variants targeting WAS gene according to a previous aspect of the invention.

The subject matter of the present invention is also further a method for inserting a transgene into the genomic WAS locus of a cell, tissue or non-human animal wherein at least one variant of the invention is introduced in said cell, tissue or non-human animal.

In a preferred embodiment, the WAS locus further allows stable expression of the transgene. In another preferred embodiment, the target sequence inside the WAS locus is only present once within the genome of said cell, tissue or individual.

In another preferred embodiment, meganuclease variants according to the present invention can be part of a kit to introduce a sequence encoding a GOI into at least one cell. In a more preferred embodiment, the at least one cell is selected form the group comprising: CHO-K1 cells; HEK293 cells; Caco2 cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; CHO-S cells; DG44 cells; K-562 cells, U-937 cells; MRC5 cells; 1MR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HT-1080 cells; HCT-1 16 cells; Hu-h7 cells; Huvec cells; Molt 4 cells.

The subject matter of the present invention is also a method for making a WAS gene knock-out or knock-in recombinant cell, comprising at least the step of:

(a) introducing into a cell, a meganuclease as defined above (I-Crel variant or single-chain derivative), so as to induce a double stranded cleavage at a site of interest of the WAS

gene comprising a DNA recognition and cleavage site for said meganuclease, simultaneously or consecutively,

(b) introducing into the cell of step (a), a targeting DNA, wherein said targeting DNA comprises (1) DNA sharing homologies to the region surrounding the cleavage site and (2) DNA which repairs the site of interest upon recombination between the targeting DNA and the chromosomal DNA, so as to generate a recombinant cell having repaired the site of interest by homologous recombination,

(c) isolating the recombinant cell of step (b), by any appropriate means.

The subject matter of the present invention is also a method for making a WAS gene knock-out or knock-in animal, comprising at least the step of:

(a) introducing into a pluripotent precursor cell or an embryo of an animal, a meganuclease as defined above, so as to induce a double stranded cleavage at a site of interest of the WAS gene comprising a DNA recognition and cleavage site for said meganuclease, simultaneously or consecutively,

(b) introducing into the animal precursor cell or embryo of step (a) a targeting DNA, wherein said targeting DNA comprises (1) DNA sharing homologies to the region surrounding the cleavage site and (2) DNA which repairs the site of interest upon recombination between the targeting DNA and the chromosomal DNA, so as to generate a genetically modified animal precursor cell or embryo having repaired the site of interest by homologous recombination,

(c) developing the genetically modified animal precursor cell or embryo of step (b) into a chimeric animal, and

(d) deriving a transgenic animal from the chimeric animal of step (c).

Preferably, step (c) comprises the introduction of the genetically modified precursor cell generated in step (b) into blastocysts so as to generate chimeric animals.

The targeting DNA is introduced into the cell under conditions appropriate for introduction of the targeting DNA into the site of interest.

For making knock-out cells/animals, the DNA which repairs the site of interest comprises sequences that inactivate the WAS gene.

For making knock-in cells/animals, the DNA which repairs the site of interest comprises the sequence of an exogenous gene of interest, and eventually a selection marker, such as the neomycin resistance gene.

In a preferred embodiment, said targeting DNA construct is inserted in a vector.

The subject matter of the present invention is also a method for making a WAS-modified cell, comprising at least the step of:

(a) introducing into a cell, a meganuclease as defined above, so as to induce a double stranded cleavage at a site of interest of the WAS gene comprising a DNA recognition and cleavage site of said meganuclease, and thereby generate genetically modified WAS gene-modified cell having repaired the double-strands break, by non-homologous end joining, and

(b) isolating the genetically modified WAS gene-modified cell of step (a), by any appropriate mean.

The subject matter of the present invention is also a method for making a WAS gene knock-out animal or animal containing a modified WAS gene, comprising at least the step of:

(a) introducing into a pluripotent precursor cell or an embryo of an animal, a meganuclease, as defined above, so as to induce a double stranded cleavage at a site of interest of the WAS gene comprising a DNA recognition and cleavage site of said meganuclease, and thereby generate genetically modified precursor cell or embryo having repaired the double-strands break by non-homologous end joining,

(b) developing the genetically modified animal precursor cell or embryo of step (a) into a chimeric animal, and

(c) deriving a transgenic animal from a chimeric animal of step (b).

Preferably, step (b) comprises the introduction of the genetically modified precursor cell obtained in step (a), into blastocysts, so as to generate chimeric animals.

The cells which are modified may be any cells of interest as long as they contain the specific target site. For making knock-in/transgenic mice, the cells are pluripotent precursor

cells such as embryo-derived stem (ES) cells, which are well-known in the art. For making recombinant human cell lines, the cells may advantageously be PerC6 (Fallaux et al., Hum. Gene Ther. 9, 1909-1917, 1998) or HEK293 (ATCC # CRL-l 573) cells.

The animal is preferably a mammal, more preferably a laboratory rodent (mice, rat, guinea-pig), or a rabbit, a cow, pig, horse or goat.

Said meganuclease can be provided directly to the cell or through an expression vector comprising the polynucleotide sequence encoding said meganuclease and suitable for its expression in the used cell.

For making recombinant cell lines expressing an heterologous protein of interest, the targeting DNA comprises a sequence encoding the product of interest (protein or RNA), and eventually a marker gene, flanked by sequences upstream and downstream the cleavage site, as defined above, so as to generate genetically modified cells having integrated the exogenous sequence of interest in the WAS gene, by homologous recombination.

The sequence of interest may be any gene coding for a certain protein/peptide of interest, included but not limited to: reporter genes, receptors, signaling molecules, transcription factors, pharmaceutically active proteins and peptides, disease causing gene products and toxins. The sequence may also encode a RNA molecule of interest including for example an interfering RNA such as ShRNA, miRNA or siRNA, well-known in the art.

The expression of the exogenous sequence may be driven, either by the endogenous WAS gene promoter or by a heterologous promoter, preferably an ubiquitous or tissue specific promoter, either constitutive or inducible, as defined above. In addition, the expression of the sequence of interest may be conditional; the expression may be induced by a site-specific recombinase such as Cre or FLP *(Akagi K, Sandig V, Vooijs M, Van der Valk M, Giovannini M, Strauss M, Berns A (May 1997). "Nucleic Acids Res. 25 (9): 1766-73.; Zhu XD, Sadowski PD (1995). J Biol Chem 270).*

Thus, the sequence of interest is inserted in an appropriate cassette that may comprise a heterologous promoter operatively linked to said gene of interest and one or more functional sequences including but not limited to (selectable) marker genes, recombinase recognition sites, polyadenylation signals, splice acceptor sequences, introns, tag for protein detection and enhancers.

The subject matter of the present invention is also a kit for making WAS gene knock¬ out or knock-in cells/animals comprising at least a meganuclease and/or one expression vector, as defined above. Preferably, the kit further comprises a targeting DNA comprising a sequence that inactivates the WAS gene flanked by sequences sharing homologies with the region of the WAS gene surrounding the DNA cleavage site of said meganuclease. In addition, for making knock-in cells/animals, the kit includes also a vector comprising a sequence of interest to be introduced in the genome of said cells/animals and eventually a selectable marker gene, as defined above.

The subject-matter of the present invention is also the use of at least one meganuclease and/or one expression vector, as defined above, for the preparation of a medicament for preventing, improving or curing a pathological condition caused by a mutation in the WAS gene as defined above, in an individual in need thereof.

The use of the meganuclease may comprise at least the step of (a) inducing in somatic tissue(s) of the donor/ individual a double stranded cleavage at a site of interest of the WAS gene comprising at least one recognition and cleavage site of said meganuclease by contacting said cleavage site with said meganuclease, and (b) introducing into said somatic tissue(s) a targeting DNA, wherein said targeting DNA comprises (1) DNA sharing homologies to the region surrounding the cleavage site and (2) DNA which repairs the WAS gene upon recombination between the targeting DNA and the chromosomal DNA, as defined above. The targeting DNA is introduced into the somatic tissues(s) under conditions appropriate for introduction of the targeting DNA into the site of interest.

According to the present invention, said double-stranded cleavage may be induced, ex vivo by introduction of said meganuclease into somatic cells from the diseased individual and then transplantation of the modified cells back into the diseased individual.

The subject-matter of the present invention is also a method for preventing, improving or curing a pathological condition caused by a mutation in the WAS gene, in an individual in need thereof, said method comprising at least the step of administering to said individual a composition as defined above, by any means. The meganuclease can be used either as a polypeptide or as a polynucleotide construct encoding said polypeptide. It is introduced into mouse cells, by any convenient means well-known to those in the art, which are appropriate for the particular cell type, alone or in association with either at least an appropriate vehicle or carrier and/or with the targeting DNA.

According to an advantageous embodiment of the uses according to the invention, the meganuclease (polypeptide) is associated with:

- liposomes, polyethyleneimine (PEI); in such a case said association is administered and therefore introduced into somatic target cells.

- membrane translocating peptides (Bonetta, The Scientist, 2002, 16, 38; Ford et al., Gene Ther., 2001, 8, 1-4 ; Wadia and Dowdy, Curr. Opin. Biotechnol., 2002, 13, 52-56); in such a case, the sequence of the variant/single-chain meganuclease is fused with the sequence of a membrane translocating peptide (fusion protein).

According to another advantageous embodiment of the uses according to the invention, the meganuclease (polynucleotide encoding said meganuclease) and/or the targeting DNA is inserted in a vector. Vectors comprising targeting DNA and/or nucleic acid encoding a meganuclease can be introduced into a cell by a variety of methods (e.g., injection, direct uptake, projectile bombardment, liposomes, electroporation). Meganucleases can be stably or transiently expressed into cells using expression vectors. Techniques of expression in eukaryotic cells are well known to those in the art. (See Current Protocols in Human Genetics: Chapter 12 "Vectors For Gene Therapy" & Chapter 13 "Delivery Systems for Gene Therapy"). Optionally, it may be preferable to incorporate a nuclear localization signal into the recombinant protein to be sure that it is expressed within the nucleus.

Once in a cell, the meganuclease and if present, the vector comprising targeting DNA and/or nucleic acid encoding a meganuclease are imported or translocated by the cell from the cytoplasm to the site of action in the nucleus.

Since meganucleases recognize a specific DNA sequence, any meganuclease developed in the context of human gene therapy could be used in other contexts (other organisms, other loci, use in the context of a landing pad containing the site) unrelated with gene therapy of WAS in human as long as the site is present.

For purposes of therapy, the meganucleases and a pharmaceutically acceptable excipient are administered in a therapeutically effective amount. Such a combination is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of the recipient. In the present context, an agent is physiologically significant if its presence results in a decrease in the severity of one or more

symptoms of the targeted disease and in a genome correction of the lesion or abnormality. Vectors comprising targeting DNA and/or nucleic acid encoding a meganuclease can be introduced into a cell by a variety of methods (e.g., injection, direct uptake, projectile bombardment, liposomes, electroporation). Meganucleases can be stably or transiently expressed into cells using expression vectors. Techniques of expression in eukaryotic cells are well known to those in the art. (See Current Protocols in Human Genetics: Chapter 12 "Vectors For Gene Therapy" & Chapter 13 "Delivery Systems for Gene Therapy").

In one embodiment of the uses according to the present invention, the meganuclease is substantially non-immunogenic, i.e., engender little or no adverse immunological response. A variety of methods for ameliorating or eliminating deleterious immunological reactions of this sort can be used in accordance with the invention. In a preferred embodiment, the meganuclease is substantially free of N-formyl methionine. Another way to avoid unwanted immunological reactions is to conjugate meganucleases to polyethylene glycol ("PEG") or polypropylene glycol ("PPG") (preferably of 500 to 20,000 daltons average molecular weight (MW)). Conjugation with PEG or PPG, as described by Davis et al. (US 4,179,337) for example, can provide non-immunogenic, physiologically active, water soluble endonuclease conjugates with anti-viral activity. Similar methods also using a polyethylene-polypropylene glycol copolymer are described in Saifer et al. (US 5,006,333).

The invention also concerns a prokaryotic or eukaryotic host cell that is modified by a polynucleotide or a vector as defined above, preferably an expression vector.

The invention also concerns a non-human transgenic animal or a transgenic plant, characterized in that all or a part of their cells are modified by a polynucleotide or a vector as defined above.

As used herein, a cell refers to a prokaryotic cell, such as a bacterial cell, or an eukaryotic cell, such as an animal, plant or yeast cell.

The subject matter of the present invention is also the use of at least one meganuclease variant, as defined above, as a scaffold for making other meganucleases. For example, further rounds of mutagenesis and selection/screening can be performed on said variants, for the purpose of making novel meganucleases.

The different uses of the meganuclease and the methods of using said meganuclease according to the present invention include the use of the I-CreI variant, the single-chain

chimeric meganuclease derived from said variant, the polynucleotide(s), vector, cell, transgenic plant or non-human transgenic mammal encoding said variant or single-chain chimeric meganuclease, as defined above.

Single-chain chimeric meganucleases able to cleave a DNA target from the gene of interest are derived from the variants according to the invention by methods well-known in the art (Epinat et al., Nucleic Acids Res., 2003, 31, 2952-62; Chevalier et al., Mol. Cell., 2002, 10, 895-905; Steuer et al., Chembiochem., 2004, 5, 206-13; International PCT Applications WO 03/078619, WO 2004/031346 and WO 2009/095793). Any of such methods, may be applied for constructing single-chain chimeric meganucleases derived from the variants as defined in the present invention. In particular, the invention encompasses also the I-CreI variants defined in the tables I, II and III.

The polynucleotide sequence(s) encoding the variant as defined in the present invention may be prepared by any method known by the man skilled in the art. For example, they are amplified from a cDNA template, by polymerase chain reaction with specific primers. Preferably, the codons of said cDNA are chosen to favor the expression of said protein in the desired expression system.

The recombinant vector comprising said polynucleotides may be obtained and introduced in a host cell by the well-known recombinant DNA and genetic engineering techniques.

The I-CreI variant or single-chain derivative as defined in the present invention are produced by expressing the polypeptide(s) as defined above; preferably said polypeptide(s) are expressed or co-expressed (in the case of the variant only) in a host cell or a transgenic animal/plant modified by one expression vector or two expression vectors (in the case of the variant only), under conditions suitable for the expression or co-expression of the polypeptide(s), and the variant or single-chain derivative is recovered from the host cell culture or from the transgenic animal/plant.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress,

USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (J. Abelson and M. Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols. 154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

**Definitions**

- Amino acid residues in a polypeptide sequence are designated herein according to the one-letter code, in which, for example, Q means Gin or Glutamine residue, R means Arg or Arginine residue and D means Asp or Aspartic acid residue.

- Amino acid substitution means the replacement of one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a peptide sequence is an amino acid substitution.

- Altered/enhanced/increased cleavage activity, refers to an increase in the detected level of meganuclease cleavage activity, see below, against a target DNA sequence by a second meganuclease in comparison to the activity of a first meganuclease against the target DNA sequence. Normally the second meganuclease is a variant of the first and comprise one or more substituted amino acid residues in comparison to the first meganuclease.

- Nucleotides are designated as follows: one-letter code is used for designating the base of a nucleoside: a is adenine, t is thymine, c is cytosine, and g is guanine. For the degenerated nucleotides, r represents g or a (purine nucleotides), k represents g or t, s represents g or c, w represents a or t, m represents a or c, y represents t or c (pyrimidine

nucleotides), d represents g, a or t, v represents g, a or c, b represents g, t or c, h represents a, t or c, and n represents g, a, t or c.

- by "endonuclease" is intended any wild-type or variant enzyme capable of catalyzing the hydrolysis (cleavage) of bonds between nucleic acids within of a DNA or RNA molecule, preferably a DNA molecule. Endonucleases do not cleave the DNA or RNA molecule irrespective of its sequence, but recognize and cleave the DNA or RNA molecule at specific polynucleotide sequences, further referred to as "target sequences" or "target sites" and significantly increased HR by specific meganuclease-induced DNA double-strand break (DSB) at a defined locus (Rouet et al, 1994; Choulika et al, 1995). Endonucleases can for example be a homing endonuclease (Paques et al. Curr Gen Ther. 2007 7:49-66), a chimeric Zinc-Finger nuclease (ZFN) resulting from the fusion of engineered zinc-finger domains with the catalytic domain of a restriction enzyme such as Fokl (Porteus et al. Nat Biotechnol. 2005 23:967-973) or a chemical endonuclease (Arimondo et al. Mol Cell Biol. 2006 26:324-333; Simon et al. NAR 2008 36:353 1-3538; Eisenschmidt et al. NAR 2005 33 :7039-7047; Cannata et al. PNAS 2008 105:9576-9581). In chemical endonucleases, a chemical or peptidic cleaver is conjugated either to a polymer of nucleic acids or to another DNA recognizing a specific target sequence, thereby targeting the cleavage activity to a specific sequence. Chemical endonucleases also encompass synthetic nucleases like conjugates of orthophenanthroline, a DNA cleaving molecule, and triplex-forming oligonucleotides (TFOs), known to bind specific DNA sequences (Kalish and Glazer Ann NY Acad Sci 2005 1058: 151-61). Such chemical endonucleases are comprised in the term "endonuclease" according to the present invention. In the scope of the present invention is also intended any fusion between molecules able to bind DNA specific sequences and agent/reagent/chemical able to cleave DNA or interfere with cellular proteins implicated in the DSB repair (Majumdar et al. J. Biol. Chem 2008 283, 17:1 1244-1 1252; Liu et al. NAR 2009 37:6378-6388); as a non limiting example such a fusion can be constituted by a specific DNA-sequence binding domain linked to a chemical inhibitor known to inhibit religation activity of a topoisomerase after DSB cleavage. Endonuclease can be a homing endonuclease, also known under the name of meganuclease.

Endonucleases according to the invention can also be derived from TALENs, a new class of chimeric nucleases using a Fokl catalytic domain and a DNA binding domain derived from Transcription Activator Like Effector (TALE), a family of proteins used in the infection

process by plant pathogens of the Xanthomonas genus(Boch, Scholze et al. 2009; Moscou and Bogdanove 2009; Christian, Cermak et al. 2010; Li, Huang et al. 201 1). The functional layout of a Fokl-based TALE-nuclease (TALEN) is essentially that of a ZFN, with the Zinc-finger DNA binding domain being replaced by the TALE domain. As such, DNA cleavage by a TALEN requires two DNA recognition regions flanking an unspecific central region. Endonucleases encompassed in the present invention can also be derived from TALENs. An endonuclease according to the present invention can be derived from a TALE-nuclease (TALEN), i.e., a fusion between a DNA-binding domain derived from a Transcription Activator Like Effector (TALE) and one or two catalytic domains.

By "meganuclease", is intended an endonuclease having a double-stranded DNA target sequence of 12 to 45 bp. Such homing endonucleases are well-known to the art (see e.g. Stoddard, Quarterly Reviews of Biophysics, 2006, 38:49-95). Homing endonucleases recognize a DNA target sequence and generate a single- or double-strand break. Homing endonucleases are highly specific, recognizing DNA target sites ranging from 12 to 45 base pairs (bp) in length, usually ranging from 14 to 40 bp in length. The homing endonuclease according to the invention may for example correspond to a LAGLIDADG endonuclease, to a HNH endonuclease, or to a GIY-YIG endonuclease. Said meganuclease is either a dimeric enzyme, wherein each domain is on a monomer or a monomelic enzyme comprising the two domains on a single polypeptide.

- by "meganuclease domain" is intended the region which interacts with one half of the DNA target of a meganuclease and is able to associate with the other domain of the same meganuclease which interacts with the other half of the DNA target to form a functional meganuclease able to cleave said DNA target.

- by "meganuclease variant" or "variant" it is intended a meganuclease obtained by replacement of at least one residue in the amino acid sequence of the parent meganuclease with a different amino acid.

- by "peptide linker" it is intended to mean a peptide sequence of at least 10 and preferably at least 17 amino acids which links the C-terminal amino acid residue of the first monomer to the N-terminal residue of the second monomer and which allows the two variant monomers to adopt the correct conformation for activity and which does not alter the specificity of either of the monomers for their targets. As non-limiting examples, first and

second monomers of a dimeric I-CreI protein can be fused together and are optionally connected by peptidic linkers consisting of or comprising sequences such as RM2 (SEQ ID NO: 34) or BQY (SEQ ID NO: 95).

- by "subdomain" it is intended the region of a LAGLIDADG homing endonuclease core domain which interacts with a distinct part of a homing endonuclease DNA target half-site.

- by "targeting DNA construct/minimal repair matrix/repair matrix" it is intended to mean a DNA construct comprising a first and second portions which are homologous to regions 5' and 3' of the DNA target in situ. The DNA construct also comprises a third portion positioned between the first and second portion which comprise some homology with the corresponding DNA sequence in situ or alternatively comprise no homology with the regions 5' and 3' of the DNA target in situ. Following cleavage of the DNA target, a homologous recombination event is stimulated between the genome containing the WAS gene and the repair matrix, wherein the genomic sequence containing the DNA target is replaced by the third portion of the repair matrix and a variable part of the first and second portions of the repair matrix.

- by "functional variant" is intended a variant which is able to cleave a DNA target sequence, preferably said target is a new target which is not cleaved by the parent meganuclease. For example, such variants have amino acid variation at positions contacting the DNA target sequence or interacting directly or indirectly with said DNA target.

- by "selection or selecting" it is intended to mean the isolation of one or more meganuclease variants based upon an observed specified phenotype, for instance altered cleavage activity. This selection can be of the variant in a peptide form upon which the observation is made or alternatively the selection can be of a nucleotide coding for selected meganuclease variant.

- by "screening" it is intended to mean the sequential or simultaneous selection of one or more meganuclease variant (s) which exhibits a specified phenotype such as altered cleavage activity.

- by "derived from" it is intended to mean a meganuclease variant which is created from a parent meganuclease and hence the peptide sequence of the meganuclease variant is

related to (primary sequence level) but derived from (mutations) the sequence peptide sequence of the parent meganuclease.

- by "I-CreI" is intended the wild-type I-CreI having the sequence of pdb accession code lg9y, corresponding to the sequence SEQ ID NO: 1 in the sequence listing.

- by "I-CreI variant with novel specificity" is intended a variant having a pattern of cleaved targets different from that of the parent meganuclease. The terms "novel specificity", "modified specificity", "novel cleavage specificity", "novel substrate specificity" which are equivalent and used indifferently, refer to the specificity of the variant towards the nucleotides of the DNA target sequence. In the present Patent Application all the I-CreI variants described comprise an additional Alanine after the first Methionine of the wild type I-CreI sequence. These variants also comprise two additional Alanine residues and an Aspartic Acid residue after the final Proline of the wild type I-CreI sequence (SEQ ID NO: 70). These additional residues do not affect the properties of the enzyme and to avoid confusion these additional residues do not affect the numeration of the residues in I-CreI or a variant referred in the present Patent Application, as these references exclusively refer to residues of the wild type I-CreI enzyme (SEQ ID NO: 1) as present in the variant, so for instance residue 2 of I-CreI is in fact residue 3 of a variant which comprises an additional Alanine after the first Methionine.

- by "I-CreI site" is intended a 22 to 24 bp double-stranded DNA sequence which is cleaved by I-CreI. I-CreI sites include the wild-type non-palindromic I-CreI homing site and the derived palindromic sequences such as the sequence 5'- t_i2C_1 1a_1oa.9a_8a_7c_6g_5t_4c_3g_2t_ ia+ ic+2g+3a+4c+5g+6t+7t+8t+9t+iog+iia+i2 (SEQ ID NO: 2), also called C1221 (Figures 3, 5 and 7).

- by "domain" or "core domain" is intended the "LAGLIDADG homing endonuclease core domain" which is the characteristic $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ fold of the homing endonucleases of the LAGLIDADG family, corresponding to a sequence of about one hundred amino acid residues. Said domain comprises four beta-strands ($\beta_1\beta_2\beta_3\beta_4$) folded in an anti-parallel beta-sheet which interacts with one half of the DNA target. This domain is able to associate with another LAGLIDADG homing endonuclease core domain that interacts with the other half of the DNA target to form a functional endonuclease able to cleave said DNA target. For example, in the case of the dimeric homing endonuclease I-CreI (163 amino acids), the LAGLIDADG homing endonuclease core domain corresponds to the residues 6 to 94.

- by "chimeric DNA target" or "hybrid DNA target" it is intended the fusion of a different half of two parent meganuclease target sequences. In addition at least one half of said target may comprise the combination of nucleotides which are bound by at least two separate subdomains (combined DNA target).

- by "beta-hairpin" is intended two consecutive beta-strands of the antiparallel beta-sheet of a LAGLIDADG homing endonuclease core domain ($\beta_1\beta_2$ or $\beta_3\beta_4$) which are connected by a loop or a turn,

- by "single-chain meganuclease", "single-chain chimeric meganuclease", "single-chain meganuclease derivative", "single-chain chimeric meganuclease derivative" or "single-chain derivative" is intended a meganuclease comprising two LAGLIDADG homing endonuclease domains or core domains linked by a peptidic spacer. The single-chain meganuclease is able to cleave a chimeric DNA target sequence comprising one different half of each parent meganuclease target sequence.

- by "DNA target", "DNA target sequence", "target sequence" , "target-site", "target" , "site", "site of interest", "recognition site", "recognition sequence", "homing recognition site", "homing site", "cleavage site" is intended a 20 to 24 bp double-stranded palindromic, partially palindromic (pseudo-palindromic) or non-palindromic polynucleotide sequence that is recognized and cleaved by a LAGLIDADG homing endonuclease such as I-Crel, or a variant, or a single-chain chimeric meganuclease derived from I-Crel. These terms refer to a distinct DNA location, preferably a genomic location, at which a double stranded break (cleavage) is to be induced by the meganuclease. The DNA target is defined by the 5' to 3' sequence of one strand of the double-stranded polynucleotide, as indicate above for CI221 . Cleavage of the DNA target occurs at the nucleotides at positions +2 and -2, respectively for the sense and the antisense strand. Unless otherwise indicated, the position at which cleavage of the DNA target by an I-Cre I meganuclease variant occurs corresponds to the cleavage site on the sense strand of the DNA target.

- by "DNA target half-site", "half cleavage site" or half-site" is intended the portion of the DNA target which is bound by each LAGLIDADG homing endonuclease core domain.

- by "chimeric DNA target" or "hybrid DNA target" is intended the fusion of different halves of two parent meganuclease target sequences. In addition at least one half of said target may comprise the combination of nucleotides which are bound by at least two separate subdomains (combined DNA target).

- by "gene" is intended the basic unit of heredity, consisting of a segment of DNA arranged in a linear manner along a chromosome, which encodes for a specific protein or segment of protein. A gene typically includes a promoter, a 5' untranslated region, one or more coding sequences (exons), optionally introns, a 3' untranslated region. The gene may further comprise a terminator, enhancers and/or silencers. by "gene" is also intended one or several part of this gene, as listed above.

- by "WAS gene", is preferably intended a WAS gene of a vertebrate or part of it, more preferably the WAS gene or part of it of a mammal such as human. WAS gene sequences are available in sequence databases, such as the NCBI/GenBank database. This gene has been described in databanks as NC000023 entry (NCBI).

- by "DNA target sequence from the WAS gene", "genomic DNA target sequence", " genomic DNA cleavage site", "genomic DNA target" or "genomic target" is intended a 22 to 24 bp sequence of the WAS gene as defined above, which is recognized and cleaved by a meganuclease variant or a single-chain chimeric meganuclease derivative.

- by "parent meganuclease" it is intended to mean a wild type meganuclease or a variant of such a wild type meganuclease with identical properties or alternatively a meganuclease with some altered characteristic in comparison to a wild type version of the same meganuclease. In the present invention the parent meganuclease can refer to the initial meganuclease from which a series of variants are derived from.

- by "vector" is intended a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked.

- by "homologous" is intended a sequence with enough identity to another one to lead to homologous recombination between sequences, more particularly having at least 95 % identity, preferably 97 % identity and more preferably 99 % or 99.5%.

- "identity" refers to sequence identity between two nucleic acid molecules or polypeptides. Identity can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base, then the molecules are identical at that position. A degree of similarity or identity between nucleic acid or amino acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. Various alignment algorithms and/or programs may be used to calculate the identity between two sequences, including FASTA, or BLAST which are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default setting.

- by "mutation" is intended the substitution, deletion, insertion of one, two, three, four, five, six, ten or more nucleotides/amino acids in a polynucleotide (cDNA, gene) or a polypeptide sequence. Said mutation can affect the coding sequence of a gene or its regulatory sequence. It may also affect the structure of the genomic sequence or the structure/stability of the encoded mRNA.

- "gene of interest" or "GOI" refers to any nucleotide sequence encoding a known or putative gene product.

- As used herein, the term "locus" is the specific physical location of a DNA sequence (e.g., of a gene) on a chromosome. The term "locus" usually refers to the specific physical location of an endonuclease's target sequence on a chromosome. Such a locus, which comprises a target sequence that is recognized and cleaved by an endonuclease according to the invention, is referred to as "locus according to the invention".

- by "safe harbor" locus of the genome of a cell, tissue or individual, is intended a gene locus wherein a transgene could be safely inserted, the disruption or deletion of said locus consecutively to the insertion not modifying expression of genes located outside of said locus.

- As used herein, the term "transgene" refers to a sequence encoding a polypeptide. Preferably, the polypeptide encoded by the transgene is either not expressed, or expressed but not biologically active, in the cell, tissue or individual in which the transgene is inserted. Most preferably, the transgene encodes a therapeutic polypeptide useful for

the treatment of an individual.

The above written description of the invention provides a manner and process of making and using it such that any person skilled in this art is enabled to make and use the same, this enablement being provided in particular for the subject matter of the appended claims, which make up a part of the original description.

As used above, the phrases "selected from the group consisting of," "chosen from," and the like include mixtures of the specified materials.

Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and subranges within a numerical limit or range are specifically included as if explicitly written out.

The above description is presented to enable a person skilled in the art to make and use the invention, and is provided in the context of a particular application and its requirements. Various modifications to the preferred embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the invention. Thus, this invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

The following non-limiting examples illustrate some aspects of the invention.

## EXAMPLES

Example 1: Engineering meganucleases targeting the WAS4 site

a) Construction of variants targeting the WAS4 site

**WAS4** is an example of a target for which meganuclease variants have been generated. The **WAS4** target sequence or WAS4.1 **(TC-CAA-AC-CTC-TCAC-AAA-AG-TGT-AT,** SEQ ID NO: 8) is located in **intron 2** of **WAS** gene at positions **1518** to **1541** of **NC000023 entry.**

The **WAS4** sequence is partially a combination of the **10CAA_P** (SEQ ID NO: 4), **5CTC_P** (SEQ ID NO: 5), **10ACA_P** (SEQ ID NO: 6) and **5TTT_P** (SEQ ID NO: 7) target sequences which are shown on **Figure 3.** These sequences are cleaved by meganucleases obtained as described in International PCT applications WO 2006/097784 and WO 2006/097853, Arnould et al. (J. Mol. Biol., 2006, 355, 443-458) and Smith et al. (Nucleic Acids Res., 2006).

Two palindromic targets, **WAS4.3** (SEQ ID NO: 10) and **WAS4.4** (SEQ ID NO: 11), and two pseudo palindromic targets, **WAS4.5** (SEQ ID NO: 12) and **WAS4.6** (SEQ ID NO: 13), were derived from **WAS4.1** (SEQ ID NO: 8) and **WAS4.2** (SEQ ID NO: 9) **(Figure 3).** Since **WAS4.3** (SEQ ID NO: 10) and **WAS4.4** (SEQ ID NO: 11) are palindromic, they are cleaved by homodimeric proteins. Therefore, homodimeric *I-CreI* variants cleaving either the **WAS4.3** palindromic target sequence of SEQ ID NO: 10 or the **WAS4.4** palindromic target sequence of SEQ ID NO: 11 were constructed using methods derived from those described in Chames *et al.* (Nucleic Acids Res., 2005, 33, e178), Arnould *et al.* (J. Mol. Biol., 2006, 355, 443.458), Smith *et al.* (Nucleic Acids Res., 2006, 34, e149) and Arnould *et al.* (Arnould *et al.* J Mol Biol. 2007 371 :49-65).

Single chain obligate heterodimer constructs were generated for the I-CreI variants able to cleave the **WAS4** target sequences when forming heterodimers. These single chain constructs were engineered using the linker RM2 (AAGGSDKYNQALSKYNQALSKYNQALSGGGGS) (SEQ ID NO: 34). During this design step, mutations K7E, K96E were introduced into the mutant cleaving **WAS4.3** (monomer 1) and mutations E8K, G19S,E61R into the mutant cleaving **WAS4.4** (monomer 2) to create the

single chain molecules: monomerl(K7E K96E)-RM2-monomer2(E8K G19S E61R) that is called SCOH- **WAS4 (Table I).**

Four additional amino-acid substitutions have been found in previous studies to enhance the activity of I-Crel derivatives: these mutations correspond to the replacement of Phenylalanine 54 with Leucine (F54L), Glutamic acid 80 with Lysine (E80K), Valine 105 with Alanine (VI 05A) and Isoleucine 132 with Valine (II 32V). Some combinations were introduced into the coding sequence of N-terminal and C-terminal protein fragment, and some of the resulting proteins were assayed for their ability to induce cleavage of the **WAS4** target.

| Single-chain encoding plasmid (SCOH-WAS4) | N-terminal mutations in Single Chains (SC) | C-terminal mutations in Single Chains (SC) | Cleavage in CHO | SC SEQ ID NO: |
|---|---|---|---|---|
| pCLS3662 (SEQ ID : 51) | 7E17A33C40Q44K68A96E132N | 8K19S30R38E44K61R68Y70S77T132V | + | 35 |
| pCLS3702 (SEQ ID : 52) | 7E17A33C40Q44K68A80K96E132V | 8K19S30R38E44K61R68Y70S77T132V | + | 36 |
| pCLS3703 (SEQ ID : 53) | 7E17A33C40Q44K68A80K96E132V | 8K19S30R38E44K61R68Y70S77T105A132V | + | 37 |

**Table I: example of SCOH- WAS4 useful for WAS4 targeting**

b) Validation of some SCOH-WAS4 variants in a mammalian cells extrachromosomal assay.

The activity of the single chain molecules against the **WAS4** target was monitored using the described CHO assay along with our internal control SCOH-RAG and I-Sce I meganucleases. All comparisons were done from 0.02 to 25ng transfected variant DNA **(Figure 4).** All the single molecules displayed **WAS4** target cleavage activity in CHO assay are listed in **Table I.** Variants shared specific behavior upon assayed dose depending on the mutation profile they bear **(Figure 4).** For example, **pCLS3702 and pCLS3703** have a similar profile than our standard control SCOH-RAG **(pCLS2222).** They are highly active at low doses, reach a maxima and decrease with increasing DNA doses. **pCLS3662** is less active than our 1-Sce I at low dose but displays a higher and more stable activity at high transfected

DNA quantity. All of the variants described are strongly active and can be used for targeting genes into the **WAS4** locus.

Example 2: Engineering meganucleases targeting the WAS5 site

a) Construction of variants targeting the WAS5 site

**WAS5** is an example of a target for which meganuclease variants have been generated. The **WAS5** target sequence or WAS5.1 **(CC-GGG-CC-CTC-GTGC-AGG-AG-AAG-AT,** SEQ ID NO: 18) is located in **exon 4** of **WAS** gene at positions **1965** to **1988** of **NC000023 entry.**

The **WAS5** sequence is partially a combination of the **10GGG_P** (SEQ ID NO: 14), **5CTC_P** (SEQ ID NO: 15), **10CTT_P** (SEQ ID NO: 16) and **5CCT_P** (SEQ ID NO: 17) target sequences which are shown on **Figure 5.** These sequences are cleaved by meganucleases obtained as described in International PCT applications WO 2006/097784 and WO 2006/097853, Arnould et al. (J. Mol. Biol., 2006, 355, 443-458) and Smith et al. (Nucleic Acids Res., 2006).

Two palindromic targets, **WAS5.3** (SEQ ID NO: 20) and **WAS5.4** (SEQ ID NO: 21), and two pseudo palindromic targets, **WAS5.5** (SEQ ID NO: 22) and **WAS5.6** (SEQ ID NO: 23), were derived from **WAS5.1** (SEQ ID NO: 18) and **WAS5.2** (SEQ ID NO: 19) **(Figure 5).** Since **WAS5.3** (SEQ ID NO: 20) and **WAS5.4** (SEQ ID NO: 21) are palindromic, they are cleaved by homodimeric proteins. Therefore, homodimeric *I-CreI* variants cleaving either the **WAS5.3** palindromic target sequence of SEQ ID NO: 20 or the **WAS5.4** palindromic target sequence of SEQ ID NO: 21 were constructed using methods derived from those described in Chames *et al.* (Nucleic Acids Res., 2005, 33, e178), Arnould *et al.* (J. Mol. Biol., 2006, 355, 443-458), Smith *et al.* (Nucleic Acids Res., 2006, 34, e149) and Arnould *et al.* (Arnould *et al.* J Mol Biol. 2007 371:49-65).

Single chain obligate heterodimer constructs were generated for the I-CreI variants able to cleave the **WAS5** target sequences when forming heterodimers. These single chain constructs were engineered using the linker RM2 (AAGGSDKYNQALSKYNQALSKYNQALSGGGGS) (SEQ ID NO: 34). During this design step, mutations K7E, K96E were introduced into the mutant cleaving **WAS5.3** (monomer 1) and mutations E8K, G19S,E61R into the mutant cleaving **WAS5.4** (monomer 2) to create the

single chain molecules: monomer 1(K7E K96E)-RM2-monomer2(E8K G19S E61R) that is called SCOH- **WAS5 (Table II).**

Four additional amino-acid substitutions have been found in previous studies to enhance the activity of I-CreI derivatives: these mutations correspond to the replacement of Phenylalanine 54 with Leucine (F54L), Glutamic acid 80 with Lysine (E80K), Valine 105 with Alanine (VI 05A) and Isoleucine 132 with Valine (1132V). Some combinations were introduced into the coding sequence of N-terminal and C-terminal protein fragment, and some of the resulting proteins were assayed for their ability to induce cleavage of the **WAS5** target.

| Single-chain encoding plasmid (SCOH-WAS5) | N-terminal mutations in Single Chains (SC) | C-terminal mutations in Single Chains (SC) | Cleavage in CHO | SC SEQ ID NO: |
|---|---|---|---|---|
| pCLS3751 (SEQ ID : 54) | 7E33R38A40R44T68Y70S7 5R77Q96E | 8K19S30S33S38T44D61R6 8Y70S75K77R | + | 38 |
| pCLS3752 (SEQ ID : 55) | 7E33R38A40R44T68Y70S7 5R77Q96E132V | 8K19S30S33S38T44D61R6 8Y70S75K77R132V | + | 39 |
| pCLS3753 (SEQ ID : 56) | 7E33R38A40R44T68Y70S7 5R77Q80K96E132V | 8K19S30S33S38T44D61R6 8Y70S75K77R105A132V | + | 40 |
| pCLS3754 (SEQ ID : 57) | 7E33R38A40R44T68Y70S7 5R77Q80K96E105A132V | 8K19S30S33S38T44D61R6 8Y70S75K77R80K105A 132V | + | 41 |
| pCLS3755 (SEQ ID : 58) | 7E33R38G40R44T68Y70S7 5R77V96E | 8K19S32N33G44K61R66H 68Y70S77T | + | 42 |
| pCLS3756 (SEQ ID : 59) | 7E33R38G40R44T68Y70S7 5R77V96E132V | 8K19S32N33G44K61R66H 68Y70S77T132V | + | 43 |
| pCLS3757 (SEQ ID : 60) | 7E33R38G40R44T68Y70S7 5R77V96E105A132V | 8K19S32N33G44K61R66H 68Y70S77T80K132V | + | 44 |
| pCLS3758 (SEQ ID : 61) | 7E33R38G40R44T68Y70S7 5R77V80K96E132V | 8K19S32N33G44K61R66H 68Y70S77T105A132V | + | 45 |
| pCLS3759 (SEQ ID : 62) | 7E33R38G40R44T68Y70S7 5R77V96E132V | 8K19S30S33S38T44D61R6 8Y70S75K77R132V | + | 46 |

**Table II: example of SCOH- WAS5 useful for WAS5 targeting**

b) Validation of some SCOH-WAS5 variants in a mammalian cells extrachromosomal assay.

The activity of the single chain molecules against the **WAS5** target was monitored using the described CHO assay along with our internal control SCOH-RAG and I-Sce I meganucleases. All comparisons were done from 0.02 to 25ng transfected variant DNA **(Figure 6).** All the single molecules displayed **WAS5** target cleavage activity in CHO assay as listed in **Table II.** Variants shared specific behavior upon assayed dose depending on the mutation profile they bear **(Figure 6).** For example, all but **pCLS3752, pCLS3753 and pCLS3754** have a similar profile than our standard control SCOH-RAG **(pCLS2222).** They are highly active at low doses, reach a maxima and decrease with increasing DNA doses. **pCLS3754** is highly active at low dose, higher than our control I-Sce I, but decreases with the transfected DNA quantity. **pCLS3752** and **pCLS3753** display an intermediate profile between the previous ones. All of the variants described are strongly active and can be used for targeting genes into the **WAS5** locus.

Example 3: Engineering meganucleases targeting the WAS6 site

a) Construction of variants targeting the WAS6 site

**WAS6** is an example of a target for which meganuclease variants have been generated. The **WAS6** target sequence or WAS6.1 **(CC-GGG-CC-CTC-GTGC-AGG-AG-AAG-AT,** SEQ ID NO: 28) is located in **intron 3** of **WAS** gene at positions **1835** to **1858** of **NC000023 entry.**

The **WAS6** sequence is partially a combination of the **10GGC_P** (SEQ ID NO: 24), **5CAC_P** (SEQ ID NO: 25), **10TCC_P** (SEQ ID NO: 26) and **5ACC_P** (SEQ ID NO: 27) target sequences which are shown on **Figure 7.** These sequences are cleaved by mega-nucleases obtained as described in International PCT applications WO 2006/097784 and WO 2006/097853, Arnould et al. (J. Mol. Biol., 2006, 355, 443-458) and Smith et al. (Nucleic Acids Res., 2006).

Two palindromic targets, **WAS6.3** (SEQ ID NO: 30) and **WAS6.4** (SEQ ID NO: 31), and two pseudo palindromic targets, **WAS6.5** (SEQ ID NO: 32) and **WAS6.6** (SEQ ID NO: 33), were derived from **WAS6.1** (SEQ ID NO: 28) and **WAS6.2** (SEQ ID NO: 29) **(Figure 7).** Since **WAS6.3** (SEQ ID NO: 30) and **WAS6.4** (SEQ ID NO: 31) are palindromic, they are cleaved by homodimeric proteins. Therefore, homodimeric \-*CreI* variants cleaving either the **WAS6.3** palindromic target sequence of SEQ ID NO: 30 or the **WAS6.4** palindromic

target sequence of SEQ ID NO: 31 were constructed using methods derived from those described in Chames *et al.* (Nucleic Acids Res., 2005, 33, e178), Arnould *et al.* (J. Mol. Biol., 2006, 355, 443-458), Smith *et al.* (Nucleic Acids Res., 2006, 34, e149) and Arnould *et al.* (Arnould *et al.* J Mol Biol. 2007 371 :49-65).

Single chain obligate heterodimer constructs were generated for the I-CreI variants able to cleave the **WAS6** target sequences when forming heterodimers. These single chain constructs were engineered using the linker RM2 (AAGGSDKYNQALSKYNQALSKYNQALSGGGGS) (SEQ ID NO: 34). During this design step, mutations K7E, K96E were introduced into the mutant cleaving **WAS6.3** (monomer 1) and mutations E8K, G19S,E61R into the mutant cleaving **WAS6.4** (monomer 2) to create the single chain molecules: monomer 1(K7E K96E)-RM2-monomer2(E8K G19S E61R) that is called SCOH- **WAS6 (Table III).**

Four additional amino-acid substitutions have been found in previous studies to enhance the activity of I-CreI derivatives: these mutations correspond to the replacement of Phenylalanine 54 with Leucine (F54L), Glutamic acid 80 with Lysine (E80K), Valine 105 with Alanine (V105A) and Isoleucine 132 with Valine (I132V). Some combinations were introduced into the coding sequence of N-terminal and C-terminal protein fragment, and some of the resulting proteins were assayed for their ability to induce cleavage of the **WAS6** target.

| Single-chain encoding plasmid (SCOH-WAS6) | N-terminal mutations in Single Chains (SC) | C-terminal mutations in Single Chains (SC) | Cleavage in CHO | SC SEQ ID NO: |
|---|---|---|---|---|
| pCLS5133 (SEQ ID : 63) | 7E30R33C44A50R68Y70S75Y77K96E105A129A132V140S | 8K19S24V28R33A38Y40Q44K54V61R68Y70S75Y77Q115T132V | + | 47 |
| pCLS5134 (SEQ ID : 64) | 7E30R33C44A50R68Y69G70S75Y77K96E105A132V129A | 8K19S24V28R33A38Y40Q44K54V61R68Y70S75Y77Q115T132V | + | 48 |
| pCLS5135 (SEQ ID : 65) | 7E19S30R33C44A50R68Y70S75Y77K96E105A129A132V140S | 8K24V28R33A38Y40Q44K61R68Y70S75Y77Q129A132V | + | 49 |
| pCLS5136 (SEQ ID : 66) | 7E19S30R33C44A50R68Y70S75Y77K96E105A129A132V140S | 8K24V28R33A38Y40Q44K61R68Y70S75Y77Q129A132V146G | + | 50 |

**Table III: example of SCOH- WAS6 useful for WAS6 targeting**

Example 4 : Cloning and extrachromosomal assay in mammalian cells.

**a) Cloning of WAS4. WAS5 and WAS6 targets in a vector for CHO screen**

The targets were cloned as follows using oligonucleotide corresponding to the target sequence flanked by gateway cloning sequence; the following oligonucleotides were ordered from PROLIGO. These oligonucleotides have the following sequences:

**WAS4:**

5'- TGGCATACAAGTTTTCCAAACCTCTCACAAAAGTGTATCAATCGTCTGTCA -3' **(SEQ ID NO: 67),**

**WAS5:**

5'- TGGCATACAAGTTTCCGGGCCCTCGTGCAGGAGAAGATCAATCGTCTGTCA -3' **(SEQ ID NO: 68),**

**WAS6:**

5'- TGGCATACAAGTTTCGGGCCTCACTTGGGGTGTGGAGACAATCGTCTGTCA -3'

**(SEQ ID NO: 69),**

Double-stranded target DNA, generated by PCR amplification of the single stranded oligonucleotide, was cloned using the Gateway protocol (INVITROGEN) into CHO reporter vector (pCLS1058). Target was cloned and verified by sequencing (MILLEGEN).

**b) Cloning of the single chain molecules**

A series of synthetic gene assembly was ordered to Gene Cust. Synthetic genes coding for the different single chain variants targeting NANOG gene were cloned in pCLS1 853 (figure 11) using Ascl and Xhol restriction sites.

**c) Extrachromosomal assay in mammalian cells**

CHO K1 cells were transfected as described in example 1.2. 72 hours after transfection, culture medium was removed and 150µ1 of lysis/revelation buffer for β-

galactosidase liquid assay was added. After incubation at 37°C, OD was measured at 420 nm. The entire process is performed on an automated Velocityl 1 BioCel platform. Per assay, 150 ng of target vector was cotransfected with an increasing quantity of variant DNA from 0.02 or 0.8 to 25 ng. The total amount of transfected DNA was completed to 175ng (target DNA, variant DNA, carrier DNA) using an empty vector (pCLS0002).

Example 5: Detection of induced mutagenesis at the endogenous site

Genomic DNA double strand break (DSB) can be repaired by homologous recombination (HR) or Non-homologous end joining (NHEJ). If the homologous recombination can restore the genomic integrity, NHEJ is though to be an error-prone mechanism which results in small insertion or deletion (InDel) at the DSB. Therefore, the detection of the mutagenesis induced by a meganuclease at its cognate endogenous locus reflects the overall activity of this meganuclease on this particular site. Thus, meganucleases designed to cleave WAS4 and WAS5 DNA targets were analyzed for their ability to induce mutagenesis at their cognate endogenous site.

Single Chain I-Crel variants targeting respectively WAS4 and WAS5 targets were cloned in the pCLS1853 plasmid. The resulting plasmids, respectively pCLS 3702, pCLS 3756, were used for this experiment. The day previous experiment, cells from the human embryonic kidney cell line, 293-H (Invitrogen) were seeded in a 10 cm dish at density of 1.2 106 cells/dish. The following day, cells were transfected with 3 µg of an empty plasmid or a meganuclease-expressing plasmid using lipofectamine (Invitrogen). 72 hours after transfection, cells were collected and diluted (dilution 1/20) in fresh culture medium. After 7 days of culture, cells were collected and genomic DNA extracted.

200ng of genomic DNA were used to amplify the endogenous locus surrounding the meganuclease cleavage site by PCR amplification.

A DNA fragment surrounding the target was amplified specifically. The specific PCR primers couples A (SEQ ID NO: 71; 5'- TGTGTCTGAACTTCAAATCTCAAC -3'), B (SEQ ID NO: 72; 5'-AGCTGTGAGTACAGCTCCTGTTCC-3'), and C (SEQ ID NO: 73; 5'-AGATGTAAGTGATCAACCAGCCC-3 '), D (SEQ ID NO: 74; 5'-GTCTGTCTGTGGATAGATGGATTGG -3'), were used to amplified fragments surrounding respectively to WAS4 (439 bp), WAS5 (322bp).

PCR amplification was performed to obtain a fragment flanked by specific adaptator sequences (SEQ ID NO 75; 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3' and SEQ ID NO: 76 5'-CCT ATCCCCTGTGTGCCTTGGC AGTCTC AG-3') provided by the company offering sequencing service (GATC Biotech AG, Germany) on the 454 sequencing system (454 Life Sciences). An average of 18,000 sequences was obtained from pools of 2 amplicons (500 ng each). After sequencing, different samples were identified based on barcode sequences introduced in the first of the above adaptators.

Sequences were then analyzed for the presence of insertions or deletions events (InDel) in the cleavage site of each WAS target.

Results are summarized in table IV. Both meganucleases show activity at the endogenous locus. Analysis of the WAS4 locus in cells transfected with pCLS3702 shows 0.19% of Indel events (10 fold above background). The single Chain I-Crel variants targeting WAS5 shows the highest activity at its endogenous locus as 0.89% of InDel events could be detected among the PCR fragment population.

| Mn | pCLS | InDel (%) |
|---|---|---|
| WAS4 | 3702 | 0.12 (0.03) |
| WAS5 | 3756 | 0.89 (0.19) |

Table IV: mutagenesis induced by meganucleases targeting the WAS gene. Two meganucleases were engineered to cleave two different DNA sequences within the WAS gene. pCLS, plasmid identification. InDel, meganuclease-induced mutagenesis determined by deep sequencing analysis of amplicons surrounding a specific target. RH freq, frequency of meganuclease-induced homologous recombination (data have been normalized for the cell plating efficiency). Values between brackets represent the sequencing background level.

Example 6: Deposited Biological Materials

The present invention also concerns the CNCM (Collection Nationale de Cultures de Microorganismes, Institut Pasteur, Paris) deposits n° CNCM 1-4344 and CNCM 1-4345 as well as the inserts respectively encoding WAS4 and WAS5 variants (respectively SEQ ID

NO: 36 and SEQ ID NO: 46) in the plasmids deposited under the respective deposit numbers above.

Example 7: Mutagenesis and Gene targeting at the endogenous WAS 5 in human cells

Cleavage activity of engineered single-chain WAS5 meganucleases was evaluated, through their ability to stimulate NHEJ and homologous recombination at the endogenous human *WAS5* locus.

**1) Materials and methods**

a) Meganuclease expression plasmids

The single-chain meganucleases used in this example are SCOH-WAS5-bl-C (SEQ ID NO: 59) and SCOH-WAS5-bl2-C (SEQ ID NO: 62) cloned in a mammalian expression vector, resulting in plasmid pCLS3756 and pCLS3759, respectively. Full protein sequences of SCOH-WAS5-M-C (SEQ ID NO: 59) and SCOH-WAS5-M2-C (SEQ ID NO: 62), respectively, as mentioned in example 2 and Table II are given by SEQ ID NO: 85 and SEQ ID NO: 86.

b) Donor repair plasmid

The donor plasmid contains a 924 bp fragment of the *WAS5* locus (position 11395001 to 11395925 on chromosome X, NT_079573.4), generated by PCR, as the left homology arm and a 1040 bp fragment of the *WAS5* locus (position 11395930 to 11396970 on chromosome X, NT_079573.4), generated by PCR, as the right homology arm. These homologies were separated with an insertion of 29bp (SEQ ID NO: 77). The resulting repair plasmid is pCLS5057 (SEQ ID NO: 78, figure 17).

c) 293H cells transfection

Human embryonic kidney 293H cells (Invitrogen) were plated at a density of $1.2 \times 10^6$ cells per 10 cm dish in complete medium (DMEM supplemented with 2 mM L-glutamine, penicillin (100 Ul/ml), streptomycin (100 µg/ml), amphotericin B (Fongizone) (0.25 µg/ml) (Invitrogen- Life Science) and 10% FBS). The next day, cells were transfected with Lipofectamine 2000 transfection reagent (Invitrogen) according to the supplier's protocol.

Briefly, for mutagenesis experiment, cells were transfected with 5µg of total DNA containing 3µg of plasmid encoding meganuclease WAS5. For gene targeting experiment,

293H cells were co-transfected with 2µg of the donor plasmid and 3 µg of single-chain meganuclease expression vector.

d) WAS5-induced mutagenesis assay

The efficiency of the meganuclease to promote mutagenesis at its endogenous recognition site was evaluated by sequencing the DNA surrounding the meganuclease cleavage site. Two days post-transfection, genomic DNA was extracted. 200ng of genomic DNA were used to amplify (PCR amplification) the endogenous locus surrounding the meganuclease cleavage site. PCR amplification is performed to obtain a fragment flanked by specific adaptor sequences (adaptor A: 5'-CCATCTCATCCCTGCGTGTCTCCGAC-NNNN-3', SEQ ID NO: 79 and adaptor B, 5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3', SEQ ID NO: 80) provided by the company offering sequencing service (GATC Biotech AG, Germany) on the 454 sequencing system (454 Life Sciences). The primers sequences used for PCR amplification were:

WAS_F:5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-NNNNAGATGTAAGTGATCAACCAGCCCTCG -3' (SEQ ID NO: 81) and

WAS_R: 5'-CCTATCCCCTGTGTGCCTTGGC AGTCTC AG-GTCTGTCTGTGGATAGATGGATTGGGAGCC -3' (SEQ ID NO: 82).

Sequences specific to the locus are underlined. The sequence NNNN in primer F1 is a Barcode sequence (Tag) needed to link the sequence with a PCR product. The percentage of PCR fragments carrying insertion or deletion at the meganuclease cleavage site is related to the mutagenesis induced by the meganuclease through NHEJ pathway in a cell population, and therefore correlates with the meganuclease activity at its endogenous recognition site. 5000 to 10000 sequences were analyzed per condition.

e) WAS5-induced gene targeting experiment

Cells were co-transfected with mammalian expression plasmids encoding SCOH-WAS5-M-C (SEQ ID NO: 85) or SCOH-WAS5-M2-C (SEQ ID NO: 86) and the donor repair plasmid pCLS5057 (SEQ ID NO: 78; Figure 17) containing 29 bp of exogenous DNA sequence (SEQ ID NO: 77) flanked by two sequences, respectively, a 924 bp fragment of the *WAS5* locus as the left homology arm and a 1040 bp fragment of the *WAS5* locus as the right homology arm, as described above. Cleavage of the native *WAS5* locus by the meganuclease

yields a substrate for homologous recombination, which may use the donor repair plasmid. Thus, the frequency with which targeted integration occurs at the *WAS5* locus is indicative of the cleavage efficiency of the genomic WAS 5 target site.

Three days post-transfection, cells were trypsinized and plated in complete medium at 10 cells per well in 96-well plates. Two weeks later, genomic DNA extraction was performed with the ZR-96 genomic DNA kit (Zymo research) according to the supplier's protocol. The detection of targeted DNA matrix integrations was performed by specific PCR amplification using the primers WAS5_F2: 5'- TTAAGGCGCGCCGGACCGCGGC -3' (specific to the 30bp of heterologous sequence; SEQ ID NO: 83) and WAS5_R2: 5'- TTTGTACCCTTGGTGCCTTATTTACC -3' (specific to a genomic sequence located outside of the homology arm; SEQ ID NO: 84).

**2) Results**

The rate of mutagenesis induced by WAS5 meganucleases at their cognate target was quantified by measuring the ratio of PCR product carrying insertion/deletion events using a PCR-sequencing strategy as described in materials and methods section. As shown in Table V, when the cell population was transfected with the SCOH-WAS5-bl-C meganuclease expression vector (pCLS3756 encoding SEQ ID NO: 85) or with SCOH-WAS5-M2-C meganuclease expression plasmid (pCLS3759 encoding SEQ ID NO: 86), the percentage of PCR fragments carried a mutation increases to 0.99% ±0.13 (n=2) and 0.47%, respectively. Mutagenesis was extremely low (0.06%) in cells transfected with empty vector.

| MN | pCLS | InDel% |
|---|---|---|
| SCOH-WAS5-b1-C | 3756 | 0.99 (0.06) |
| SCOH-WAS5-b12-C | 3759 | 0.47 (0.06) |

**Table V: Mutagenesis-induced by meganucleases targeting the WAS gene.** Two meganucleases were engineered to cleave the same WAS5 target. pCLS: plasmid identification number. InDel meganuclease-induced mutagenesis was determined 7 days post-transfection by deep sequencing analysis of amplicons surrounding a specific target. Values between brackets represent the sequencing background level.

Gene targeting frequencies induced by the WAS5 meganucleases were quantified after co-transfection of the 293H cells with meganuclease expressing vector and donor repair plasmid. As a control for spontaneous recombination, 293H cells were also transfected with the donor repair plasmid alone. The cells were then plated at 10 cells per well in 96-well microplates. Genomic DNA derived from these cells was analyzed for gene targeting by PCR as described in Material and Methods. In the absence of meganuclease (repair plasmid alone), 2 PCR positive signals was detected among the 2880 cells analyzed in pools. In contrast, in the presence of the SCOH-WAS5-bl-C meganuclease (pCLS3756 encoding SEQ ID NO: 85), 77 positives PCR were detected among the 2880 cells analyzed indicating a frequency of recombination corrected by transfection and plating efficiency of 11.25 %. In the presence of the SCOH-WAS5-bl2-C meganuclease (pCLS3759 encoding SEQ ID NO: 86), 53 positives PCR were detected among the 2880 cells analyzed indicating a frequency of recombination corrected of 7.5 %. Results are presented in Table VI. These results demonstrate that the two single chain molecules SCOH-WAS5-bl-C and SCOH-WAS5-bl2-C (SEQ ID NO: 85 and 86) are both able to induce high levels of gene targeting at the endogenous *WAS5* locus.

| **pCLS** | **Cells per well** | **PCR+ events** | **Gene targeting frequency** |
|---|---|---|---|
| SCOH-WAS5-b1-C (pCLS3756) + Repair plasmid pCLS5057 | 10 | 77/288 | 11.25% |
| SCOH-WAS5-b12-C (pCLS3759) + Repair plasmid pCLS5057 | 10 | 53/288 | 7.5 % |
| +Repair plasmid pCLS5057 | 10 | 2/288 | 0.45% |

**Table VI: Frequency of gene targeting events at the WAS5 locus in human 293H cells**
Two meganucleases were engineered to cleave the same WAS target. pCLS, plasmid identification. Gene targeting frequency induced by meganucleases have been normalized for the transfection efficiency (72%) and for cell plating efficiency (X3).

Example 8: Making of a new WAS5 meganuclease more specific and more active.

The mutational landscape of the N-terminal mutant of the single chain WAS5 meganuclease encoded by pCLS3756 (SEQ ID NO: 85) was changed. Mutations Q44T, R70S and D75R were reverted and the new mutations R68D and I77R were introduced. The new WAS5 single chain meganuclease obtained was directly cloned into the mammalian expression vector and designed as pCLS9067 (SEQ ID NO: 93 encoding SEQ ID NO: 94). Activities of both meganucleases encoded by pCLS3756 (SEQ ID NO: 85) and pCLS9067 (SEQ ID NO: 94) were further tested using an extrachromosomal Single Strand Annealing (SSA) assay in CHO-K1 cells. Their toxicity were assessed using a cell survival assay. Finally, their ability to induce gene targeting events in 293H cells were evaluated.

**1) Material and methods**

a) Generation of the pCLS9067

Site-directed mutagenesis was performed using the pCLS3756 as template to generate the pCLS9067. Three independent PCR reactions were done using the following primers: CMVFor (5'-cgcaaatgggcggtaggcgt-3'; SEQ ID NO: 87) / Was544QRev (5'-cttttgagtcacctgaaaggtcaaacg-3'; SEQ ID NO: 88), Was544QFor (5'-cgtttgacctttcaggtgactcaaaag-3'; SEQ ID NO: 89) / Was5DRRev (5'-ttcgcttaaacggtaatcggaaacgctaccgcgatcgtctacgtaacc-3'; SEQ ID NO: 90) and Was5DRFor (5'-ggttacgtagacgatcgcggtagcgtttccgattaccgtttaagcgaa-3'; SEQ ID NO: 91) / V5Reverse(5'-cgtagaatcgagaccgaggagagg-3'; SEQ ID NO: 92). The three PCR fragments were gel purified and a PCR assembly was realized using the CMVFor / V5reverse oligonuleotides. The final PCR product was digested with the AscI and XhoI restriction enzymes and ligated into the pCLS3756 digested by these same enzymes.

b) Extrachromosomal SSA assay

CHO-K1 cells were transfected with 100 ng of DNA containing various amounts of meganuclease expression vectors pCLS3756 and pCLS9067 (0 to 25 ng) and 75 ng of the

reporter plasmid, in the presence of Polyfect transfection reagent in accordance with the manufacturer's protocol (Qiagen). The culture medium was removed 72 hours after transfection, and 150µl of lysis/detection buffer was added for β-galactosidase liquid assay (typically, for 1 liter of buffer, we used 100 ml of lysis buffer (10 mM Tris-HCl pH7.5, 150 mM NaCl, 0.1% Triton X100, 0.1 mg/ml BSA, protease inhibitors), 10ml of Mg 100 x buffer ($MgCl_2$ 100 mM, 2-mercaptoethanol 35%), 110ml of an 8 mg/ml solution of ONPG and 780 ml of 0.1 M sodium phosphate pH7.5). After incubation at 37°C, optical density at 420nm was measured. The entire process was performed on 96-well plate format using an automated Velocityl 1 BioCel platform.

c) Cell Survival assay

The CHO-K1 cell line was used to seed plates at a density of $2.5x10^3$ cells per well. The next day, varying amounts of meganuclease expression vector and a constant amount of GFP-encoding plasmid (10 ng) were used to transfect the cells with a total quantity of 200 ng DNA using Polyfect reagent. GFP positive cells were counted by flow cytometry (Guava EasyCyte, Guava Technologies) on days 1 and 6 post transfection. Cell survival is expressed as a percentage and was calculated as a ratio (meganuclease-transfected cell expressing GFP on day 6 / control-transfected cell expressing GFP on day 6) corrected for the transfection efficiency determined on day 1.

d) Gene targeting experiment

Cells were co-transfected with mammalian expression plasmid pCLS9067 (SEQ ID NO: 93) encoding SEQ ID NO: 94 single-chain meganuclease and the donor repair plasmid pCLS5057 (SEQ ID NO: 78; Figure 17) containing 29 bp of exogenous DNA sequence (SEQ ID NO: 77) flanked by two sequences, respectively, a 924 bp fragment of the *WAS5* locus as the left homology arm and a 1040 bp fragment of the *WAS5* locus as the right homology arm, as described in example 7. Cleavage of the native *WAS5* locus by the meganuclease yields a substrate for homologous recombination, which may use the donor repair plasmid. Thus, the frequency with which targeted integration occurs at the *WAS5* locus is indicative of the cleavage efficiency of the genomic WAS5 target site.

Three days post-transfection, cells were trypsinized and plated in complete medium at 10 cells per well in 96-well plates. Two weeks later, genomic DNA extraction was performed

with the ZR-96 genomic DNA kit (Zymo research) according to the supplier's protocol. The detection of targeted DNA matrix integrations was performed by specific PCR amplification using the primers WAS5_F2: 5'- TTAAGGCGCGCCGGACCGCGGC -3' (specific to the 30bp of heterologous sequence; SEQ ID NO: 83) and WAS5_R2: 5'- TTTGTACCCTTGGTGCCTTATTTACC -3' (specific to a genomic sequence located outside of the homology arm; SEQ ID NO: 84).

**2) Results**

The activity of the new WAS5 meganuclease (SEQ ID NO: 94 encoded by pCLS9067) was monitored using an extrachromosomal SSA assay. Figure 18, panel A shows that its activity profile can be superimposed to that of pCLS3756. Then, its potential toxicity was assessed using a cell survival assay. While pCLS3756 presents some signs of toxicity with a cell survival rate of 55% at 12.5 ng of transfected DNA, pCLS9067 is not toxic with a cell survival rate of 92% at the same dose of transfected DNA (Figure 18, panel B). Hence, the meganuclease encoded by pCLS9067 has a better activity / toxicity ratio than the meganuclease encoded by pCLS3756. To evaluate the potential benefit of such improvement a gene targeting experiment was performed in 293H cells. By co-transfecting into the cells the expression plasmid for the meganuclease and the repair plasmid pCLS5057 (SEQ ID NO: 78; Figure 17), we measured the frequency of incorporation of an exogenous sequence of 29 bp (SEQ ID NO: 77) at the WAS5 locus. The new WAS5 meganuclease (SEQ ID NO: 94 encoded by pCLS9067) can achieve a gene targeting frequency of 15.6 % to be compared to the 10.3 % obtained with pCLS3756. Therefore, use of single chain meganuclease (SEQ ID NO: 94) encoded by pCLS9067 (SEQ ID NO: 93) can be advantageous for genomic modifications purposes.

Modifications and other embodiments

Numerous modifications and variations on the present invention are possible in light of the above teachings. It is, therefore, to be understood that within the scope of the accompanying claims, the invention may be practiced otherwise than as specifically described herein. Various modifications and variations of the described WAS endonucleases and meganuclease products, compositions and methods as well as the concept of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the

invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed is not intended to be limited to such specific embodiments. Various modifications of the described modes for carrying out the invention which are obvious to those skilled in the medical, biological, chemical or pharmacological arts or related fields are intended to be within the scope of the following claims. Unless specifically defined herein below, all technical and scientific terms used herein have the same meaning as commonly understood by a skilled artisan in the fields of gene therapy, biochemistry, genetics, and molecular biology. All methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, with suitable methods and materials being described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. Further, the materials, methods, and examples are illustrative only and are not intended to be limiting, unless otherwise specified.

## REFERENCES


1. Adjali, O., et al. (2005 ). J Clin Invest **115**(**8**): 2287-95.
2. Aiuti, A., et al. (2002). Science **296**(5577): 241 0-3.
3. Aldrich, R. A., et al. (1954). Pediatrics **13**(2): 133-9.
4. Arnould, S., et al. (2006). J Mol Biol **355**(**3**): 443-58.
5. Arnould, S., et al. (2007). J Mol Biol **371**( **1**): 49-65.
6. Bach, F. H., et al. (1968). Lancet **2**(7583): 1364-6.
7. Bainbridge, J. W., et al. (2008). N Engl J Med **358**(2 1): 223 1-9.
8. Cavazzana-Calvo, M., et al. (2000). Science **288**(5466): 669-72.
9. Chames, P., et al. (2005). Nucleic Acids Res **33**(20): e l 78.
10. Chang, A. H. and M. Sadelain (2007). Mol Ther **15**(3): 445-56.
11. Chen, Z., et al. (2009). Protein Eng Pes Sel **22**(4): 249-56.
12. Chen, Z. and H. Zhao (2005). Nucleic Acids Res **33**( 1**8**): e l 54.
13. Chevalier, B. S. and B. L. Stoddard (2001). Nucleic Acids Res **29**(18): 3757-74.
14. Choulika, A., et al. (1995). Mol Cell Biol **15**(4): 1968-73.
15. De Luca, M., et al. (2009). "Gene therapy of inherited skin adhesion disorders: a critical overview." Br J Dermatol.
16. Doyon, J. B., et al. (2006). J Am Chem Soc **128**(**7**): 2477-84.
17. Epinat, J. C , et al. (2003). Nucleic Acids Res **3 1**( 1 1): 2952-62.
18. Galetto, R., et al. (2009). Expert Opin Biol Ther **9**(10): 1289-303.
19. Gaspar, H. B., et al. (2004). Lancet **364**(9452): 2 181-7.
20. Gatti, R. A., et al. (1968). Lancet **2**(7583): 1366-9.
21. Gimble, F. S., et al. (2003). J Mol Biol **334**(**5**): 993-1008.
22. Hacein-Bey-Abina, S., et al. (2008). J Clin Invest **118**(**9**): 3 132-42.
23. Hacein-Bey-Abina, S., et al. (2003). Science **302**(5644): 415-9.

24. Howe, S. J., et al. (2008). J Clin Invest **118**(**9**): 3143-50.

25. Kim, Y. G., et al. (1996). Proc Natl Acad Sci U S A **93**(3): 1156-60.

26. Korkegian, A., et al. (2005 ). Science **308**(5723): 857-60.

27. Lemahieu, V., et al. (1999). Hum Mutat **14**(**1**): 54-66.

28. Li, H., et al. (2009). Nucleic Acids Res.

29. Maguire, A. M., F. Simonelli, et al. (2008). N Engl J Med **358**(**2**1): 2240-8.

30. McConnell Smith, A., et al. (2009). Proc Natl Acad Sci U S A **106**(13): 5099-104.

31. Montini, E., et al. (2009). J Clin Invest **119**(4): 964-75.

32. Paques, F. and P. Duchateau (2007). Curr Gene Ther 7(1): 49-66.

33. Paques, F. and J. E. Haber (1999). Microbiol Mol Biol Rev **63**(2): 349-404.

34. Perez, E. E., et al. (2008). Nat Biotechnol **26**(7): 808-16.

35. Puchta, H., et al. (1996). Proc Natl Acad Sci U S A **93**(10): 5055-60.

36. Puck, J. M. and F. Candotti (2006). N Engl J Med **355**(17): 1759-61 .

37. Redondo, P., et al. (2008). Nature **456**(7218): 107-1 1.

38. Rouet, P., et al. (1994). Proc Natl Acad Sci U S A **9 1**(13): 6064-8.

39. Rouet, P., et al. (1994). Mol Cell Biol **14**(12): 8096-106.

40. Seligman, L. M., et al. (2002). Nucleic Acids Res **30**(17): 3870-9.

41. Smith, J., et al. (2000). Nucleic Acids Res **28**(17): 3361-9.

42. Smith, J., et al. (2006). Nucleic Acids Res. **Nucleic Acids Res. 2006** ;34**(22** ):el49. **Epub 2006 Nov 27.**

43. Stoddard, B. L. (2005). Q Rev Biophys **38**(**1**): 49-95.

44. Wiskott, A. (1937) Familiarer, angeborener Morbus Werlhofii Mschr. Kinderheilk. 68: 212-216,

45. Yee, J. K., et al. (1987). Proc Natl Acad Sci U S A **84**(15): 5197-201 .

46. Yu, S. F., et al. (1986). Proc Natl Acad Sci U S A **83**(10): 3194-8.

47. Sargent et al. Mol Cell Biol 1997 17: 267-277

48. Cohen-Tannoudji et al. Mol Cell Biol 1998 18: 1444-1448

49. Donoho et al. Mol Cell Biol 1998 18 : 4070-4078

50. Elliott et al. Mol Cell Biol 1998 18 : 93-101

51. Chevalier et al. Nat Struct Biol 2001 8 : 312-316

52. Chevalier et al. J Mol Biol 2003 329 : 253-269

53. Moure et al. J Mol Biol 2003 334 : 685-693,

54. Silva et al. J Mol Biol 1999 286 : 1123-1136

55. Bolduc et al. Genes Dev 2003 17 : 2875-2888

56. Ichiyanagi et al. J Mol Biol 2000 300 : 889-901

57. Moure et al. Nat Struct Biol 2002 9 : 764-770

58. Chevalier et al. Mol Cell 2002 10 : 895-905

59. Seligman et al. Genetics 1997 147 : 1653-1664

60. Sussman et al. J Mol Biol 2004 342 : 31-41

61. Rosen et al. Nucleic Acids Res 2006 34 : 4791-4800

62. Ashworth et al. Nature 2006 441 : 656-659

63. Argast et al. J Mol Biol 1998 280 : 345-353

64. Jurica et al. Mol Cell 1998 2 : 469-476

65. Chevalier et al. Biochemistry 2004 43 : 14015-14026

66. Prieto et al., Nucleic Acids Res., Epub 22 April 2007

67. Puchta et al. Nucleic Acids Res 1993 21 : 5034-5040

68. Boch, J., H. Scholze, et al. (2009). "Breaking the code of DNA binding specificity of TAL-type III effectors." Science **326** (5959): 1509-12.

69. Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science **244** (4910): 1288-92.

70. Christian, M., T. Cermak, et al. (2010). "Targeting DNA double-strand breaks with TAL effector nucleases." Genetics **186**(**2**): 757-61 .

71. Li, T., S. Huang, et al. (2010). "TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and Fokl DNA-cleavage domain." Nucleic Acids Res **39**(1): 359-72.

72. Moscou, M. J. and A. J. Bogdanove (2009). "A simple cipher governs DNA recognition by TAL effectors." Science **326** (5959): 1501 .

73. Smithies, O. (2001). "Forty years with homologous recombination." Nat Med 7(10): 1083-6.

## CLAIMS

1. An endonuclease that recognizes a polynucleotide sequence within the WAS gene locus.

2. The endonuclease of claim 1 that recognizes a sequence within an exon of the WAS gene.

3. The endonuclease of claim 1 that recognizes a sequence within an intron of the WAS gene.

4. The endonuclease of claim 1 that is derived from a Transcription Activator Like Effector (TALE).

5. The endonuclease of claim 4 that is a fusion between a DNA-binding domain derived from a Transcription Activator Like Effector (TALE) and one or two catalytic domains having endonuclease activity.

6. The endonuclease of claim 1 that is a meganuclease.

7. The endonuclease of claim 1 that is a meganuclease of the I-Crel family.

8. The meganuclease of claim 6 that recognizes the WAS4. 1 polynucleotide target (SEQ ID NO: 8) or a derivative of said target site.

9. The meganuclease of claim 6 that recognizes the WAS5.1 polynucleotide target (SEQ ID NO: 18) or a derivative of said target site.

10. The meganuclease of claim 6 that recognizes the WAS6.1 polynucleotide target (SEQ ID NO: 28) or a derivative of said target site.

11. The endonuclease of claim 6 which comprises a polypeptide sequence selected from any one of SEQ ID NOS: 35 to 37, SEQ ID NOS: 38 to 46 or SEQ ID NOS: 47 to 50, or a polypeptide sequence that is at least 95%, 97.5% or 99% identical thereto.

12. The meganuclease of claim 6, which is a homodimer.

13. The meganuclease of claim 6, which is a heterodimer.

14. The meganuclease variant of claim 6, which is a obligate heterodimer.

15. The meganuclease variant of claim 6, which is a single chain.

16. A pharmaceutical composition comprising the meganuclease of claim 6.

17. A polynucleotide that encodes the meganuclease of claim 6 or that encodes a fragment thereof having meganuclease activity.

18. A vector comprising the polynucleotide of claim 17.

19. A host cell containing the vector of claim 18.

20. A pharmaceutical composition comprising the polynucleotide of claim 17, optionally contained within a vector of claim 18, or contained within a host cell of claim 19.

21. A method for producing a meganuclease recognizing a target sequence within the WAS gene comprising culturing the host cell of claim 19 for a time and under conditions sufficient for expression of said meganuclease and recovering said meganuclease.

22. A method for engineering or modifying a sequence within the WAS gene comprising contacting a polynucleotide with at least one endonuclease that recognizes a target polynucleotide sequence within said WAS gene.

23. The method of claim 22, wherein said endonuclease recognizes a target in a WAS gene exon.

24. The method of claim 22, wherein said endonuclease recognizes a target in a WAS gene intron.

25. The method of claim 22, wherein said endonuclease induces a double-stranded break in a WAS gene polynucleotide.

26. The method of claim 22, further comprising contacting said polynucleotide with at least one additional endonuclease that recognizes a polynucleotide sequence at the WAS locus.

27. The method of claim 22, which comprises insertion of a gene or polynucleotide of interest at the WAS gene.

28. The method of claim 22, which comprises insertion of a functional WAS gene, functional WAS gene exon, or other WAS gene polynucleotide at the WAS gene.

29. The method of claim 22, which comprises gene correction or gene modification of a sequence at the WAS gene.

30. The method of claim 22 wherein said endonuclease is derived from a Transcription Activator Like Effector (TALE).

31. The method of claim 22 wherein said endonuclease is a fusion between a DNA-binding domain derived from a Transcription Activator Like Effector (TALE) and one or two catalytic domains having endonuclease activity.

32. The method of claim 22, wherein said endonuclease is a meganuclease.

33. The method of claim 22, wherein said endonuclease is a meganuclease of the I-Crel family.

34. The method of claim 23, wherein said endonuclease recognizes a target polynucleotide sequence selected from the group consisting of WAS4.1 (SEQ ID NO: 8), WAS 5 (SEQ ID NO: 18), WAS6 (SEQ ID NO: 28) and derivatives thereof.

35. A transgenic cell into which a control, expression of modulator polynucleotide has been operatively inserted using the endonuclease of claim 1 so as to modulate the expression of the WAS gene.

36. A deposit of biological material containing meganuclease variants having the CNCM accession number CNCM 1-4344 or CNCM 1-4345.

37. DNA inserts of SEQ ID NO: 36 and SEQ ID NO: 46 which are part of the biological material deposits of claim 36.

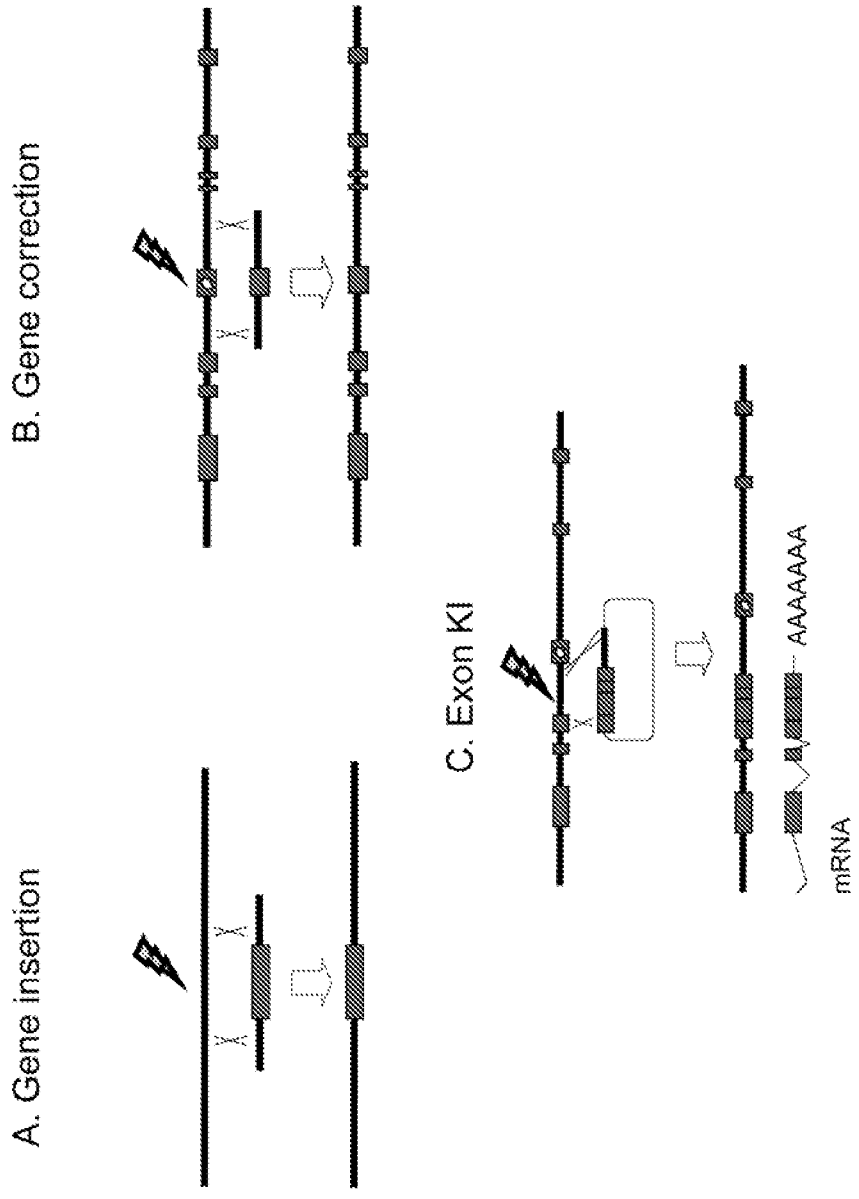


Figure 1

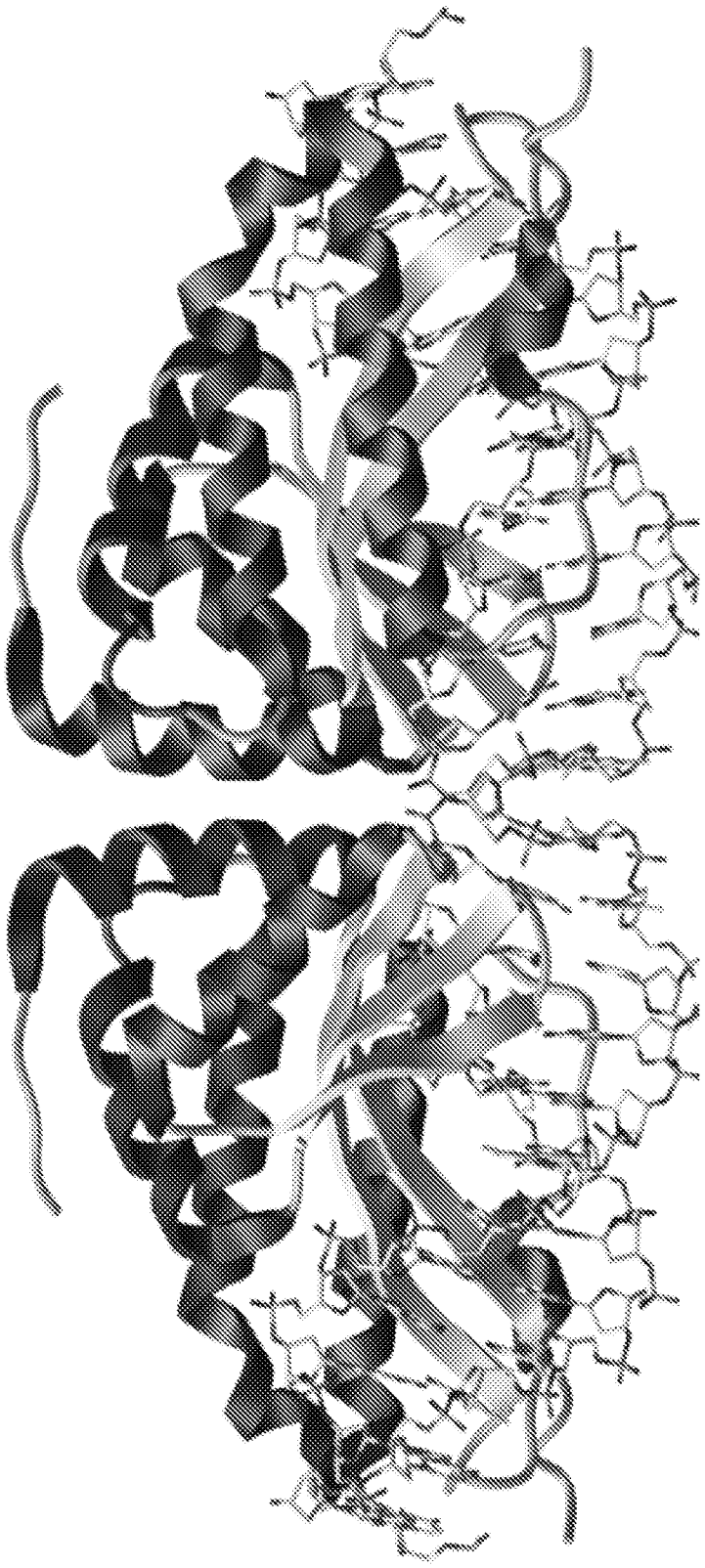

Figure 2a

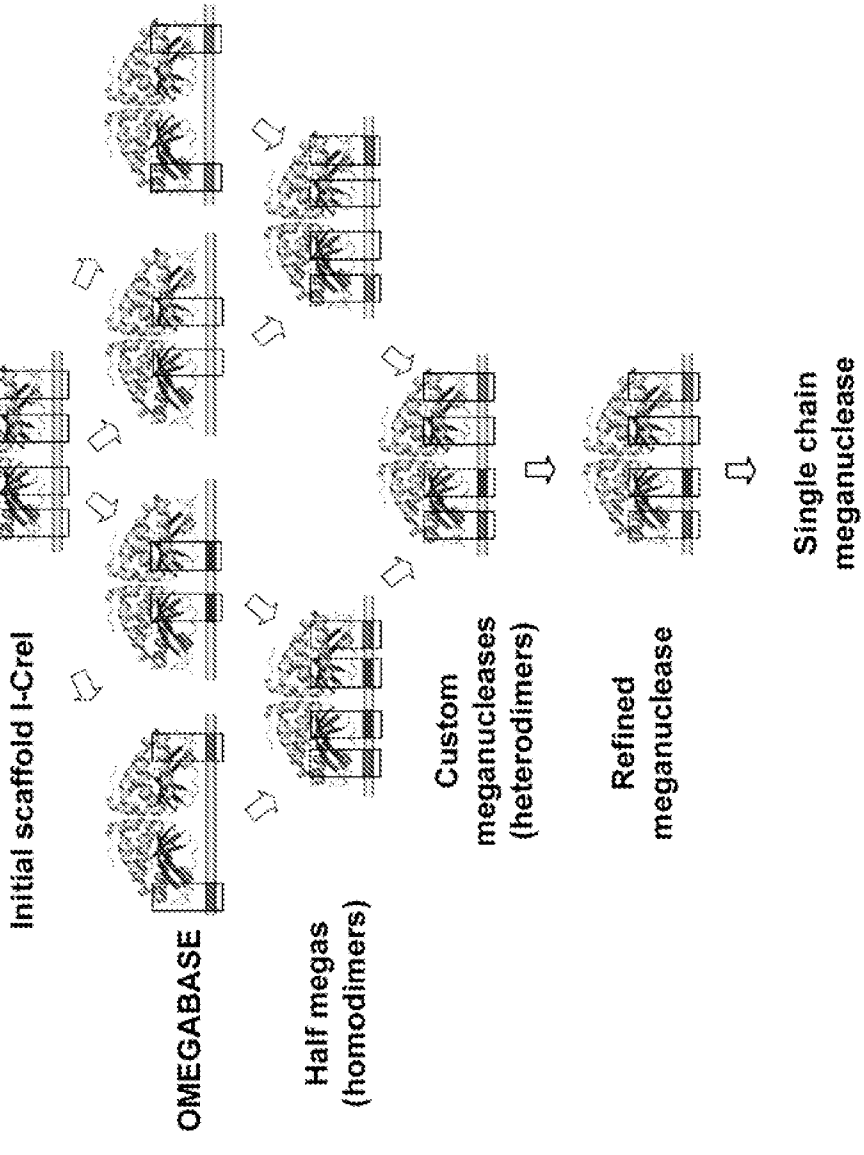
Initial scaffold I-CreI
OMEGABASE
Half megas (homodimers)
Custom meganucleases (heterodimers)
Refined meganuclease
Single chain meganuclease

Figure 2b

| | | |
|---|---|---|
| C1221 | TC-AAA-AC-GTC-GTAC-GAC-GT-TTT-GA | Seq ID N° 2 |
| 10CAA_P | TC-CAA-AC-GTC-GTAC-GAC-GT-TTG-GA | Seq ID N° 4 |
| 5CTC_P | TC-AAA-AC-CTC-GTAC-GAG-GT-TTT-GA | Seq ID N° 5 |
| 10ACA_P | TC-ACA-AC-GTC-GTAC-GAC-GT-TGT-GA | Seq ID N° 6 |
| 5TTT_P | TC-AAA-AC-TTT-GTAC-AAA-GT-TTT-GA | Seq ID N° 7 |
| WAS4.1 | TC-CAA-AC-CTC-TCAC-AAA-AG-TGT-AT | Seq ID N° 8 |
| WAS4.2 | TC-CAA-AC-CTC-GTAC-AAA-AG-TGT-AT | Seq ID N° 9 |
| WAS4.3 | TC-CAA-AC-CTC-GTAC-GAG-GT-TTG-GA | Seq ID N° 10 |
| WAS4.4 | AT-ACA-CT-TTT-GTAC-AAA-AG-TGT-AT | Seq ID N° 11 |
| WAS4.5 | TC-CAA-AC-CTC-TCAC-GAG-GT-TTG-GA | Seq ID N° 12 |
| WAS4.6 | AT-ACA-CT-TTT-TCAC-AAA-AG-TGT-AT | Seq ID N° 13 |

Figure 3

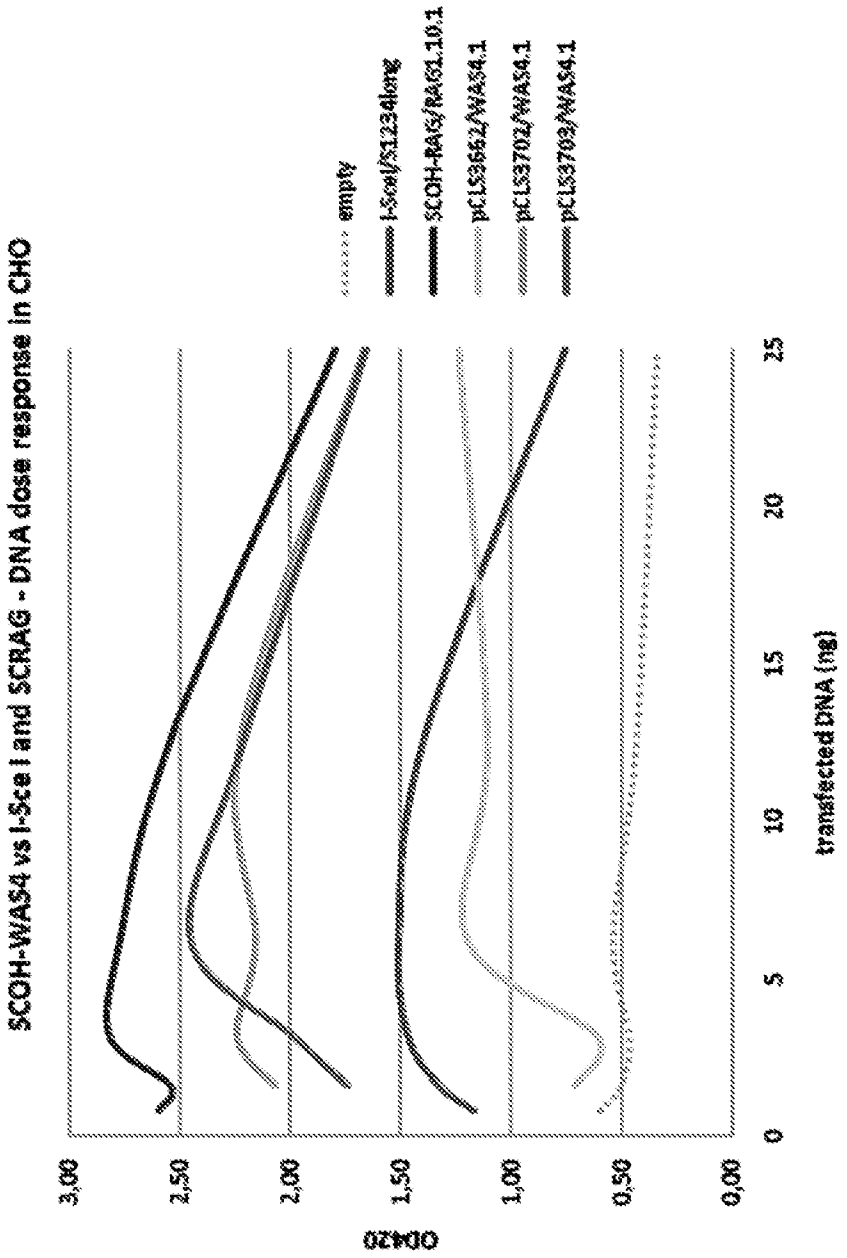
SCOH-WAS4 vs I-Sce I and SCRAG - DNA dose response in CHO
empty
I-SceI/S1234long
SCOH-RAG/RAG1.10.1
pCLS3662/WAS4.1
pCLS3702/WAS4.1
pCLS3703/WAS4.1
OD420
3,00
2,50
2,00
1,50
1,00
0,50
0,00
0
5
10
15
20
25
transfected DNA (ng)

Figure 4

| | | |
|---|---|---|
| C1221 | TC-AAA-AC-GTC-GTAC-GAC-GT-TTT-GA | Seq ID N° 2 |
| 10GGG_P | TC-GGG-AC-GTC-GTAC-GAC-GT-CCC-GA | Seq ID N° 14 |
| 5CTC_P | TC-AAA-AC-CTC-GTAC-GAG-GT-TTT-GA | Seq ID N° 15 |
| 10CTT_P | TC-CTT-AC-GTC-GTAC-GAC-GT-AAG-GA | Seq ID N° 16 |
| 5CCT_P | TC-AAA-AC-CCT-GTAC-AGG-GT-TTT-GA | Seq ID N° 17 |
| WAS5.1 | CC-GGG-CC-CTC-GTGC-AGG-AG-AAG-AT | Seq ID N° 18 |
| WAS5.2 | CC-GGG-CC-CTC-GTAC-AGG-AG-AAG-AT | Seq ID N° 19 |
| WAS5.3 | CC-GGG-CC-CTC-GTAC-GAG-GG-CCC-GG | Seq ID N° 20 |
| WAS5.4 | AT-CTT-CT-CCT-GTAC-AGG-AG-AAG-AT | Seq ID N° 21 |
| WAS5.5 | CC-GGG-CC-CTC-GTGC-GAG-GG-CCC-GG | Seq ID N° 22 |
| WAS5.6 | AT-CTT-CT-CCT-GTGC-AGG-AG-AAG-AT | Seq ID N° 23 |

Figure 5

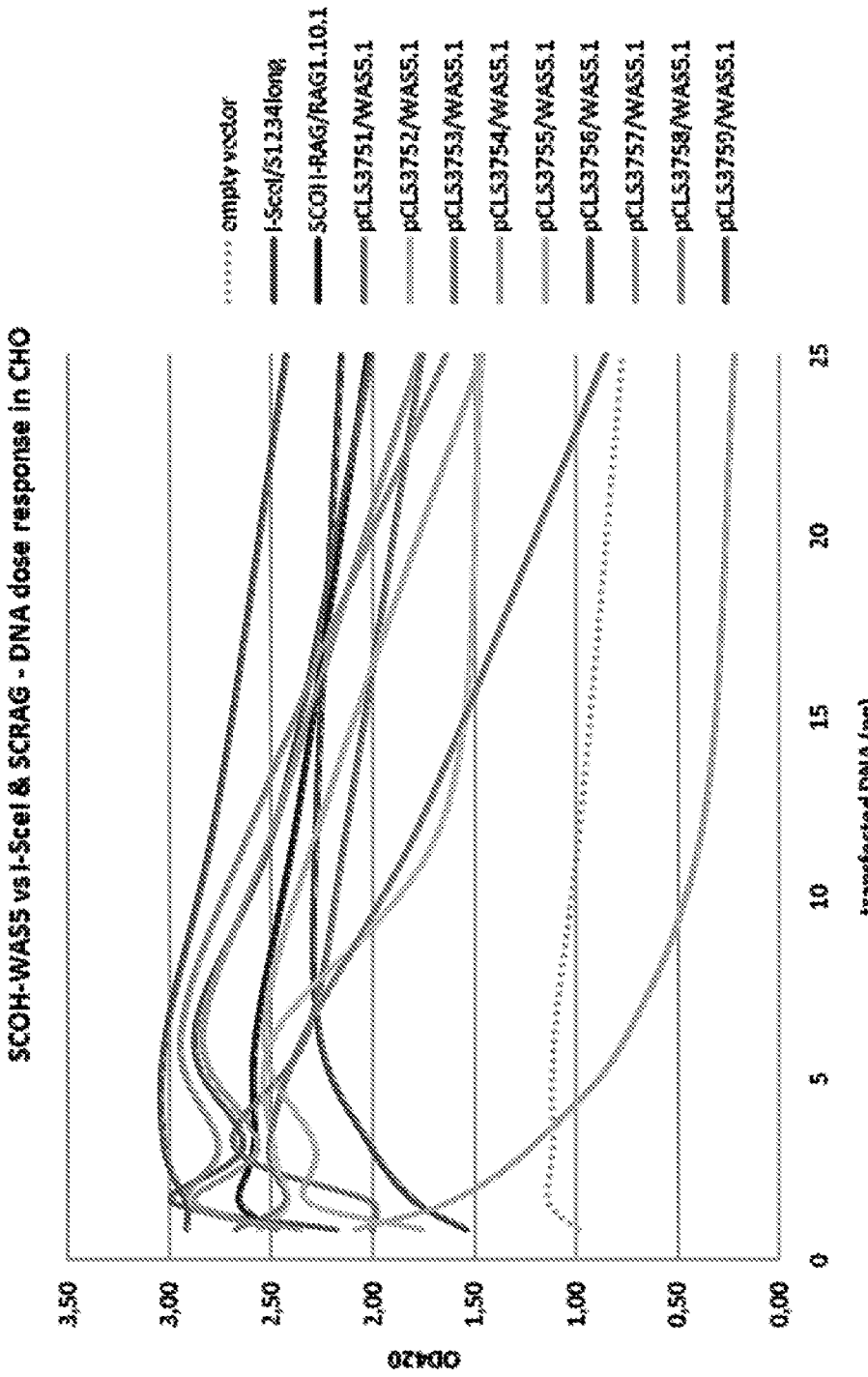
SCOH-WAS5 vs I-SceI & SCRAG - DNA dose response in CHO
empty vector
I-SceI/S1234long
SCOH-RAG/RAG1.10.1
pCL53751/WAS5.1
pCL53752/WAS5.1
pCL53753/WAS5.1
pCL53754/WAS5.1
pCL53755/WAS5.1
pCL53756/WAS5.1
pCL53757/WAS5.1
pCL53758/WAS5.1
pCL53759/WAS5.1
OD420
3,50
3,00
2,50
2,00
1,50
1,00
0,50
0,00
0
5
10
15
20
25
transfected DNA (ng)

Figure 6

| | | |
|---|---|---|
| C1221 | TC-AAA-AC-GTC-GTAC-GAC-GT-TTT-GA | Seq ID N° 2 |
| 10GGC_P | TC-GGC-AC-GTC-GTAC-GAC-GT-GCC-GA | Seq ID N° 24 |
| 5CAC_P | TC-AAA-AC-CAC-GTAC-GTG-GT-TTT-GA | Seq ID N° 25 |
| 10TCC_P | TC-TCC-AC-GTC-GTAC-GAC-GT-GGA-GA | Seq ID N° 26 |
| 5ACC_P | TC-AAA-AC-ACC-GTAC-GGT-GT-TTT-GA | Seq ID N° 27 |
| WAS6.1 | CG-GGC-CT-CAC-TTGG-GGT-GT-GGA-GA | Seq ID N° 28 |
| WAS6.2 | CG-GGC-CT-CAC-GTAC-GGT-GT-GGA-GA | Seq ID N° 29 |
| WAS6.3 | CG-GGC-CT-CAC-GTAC-GTG-AG-GCC-CG | Seq ID N° 30 |
| WAS6.4 | TC-TCC-AC-ACC-GTAC-GGT-GT-GGA-GA | Seq ID N° 31 |
| WAS6.5 | CG-GGC-CT-CAC-TTGG-GTG-AG-GCC-CG | Seq ID N° 32 |
| WAS6.6 | TC-TCC-AC-ACC-TTGG-GGT-GT-GGA-GA | Seq ID N° 33 |

Figure 7

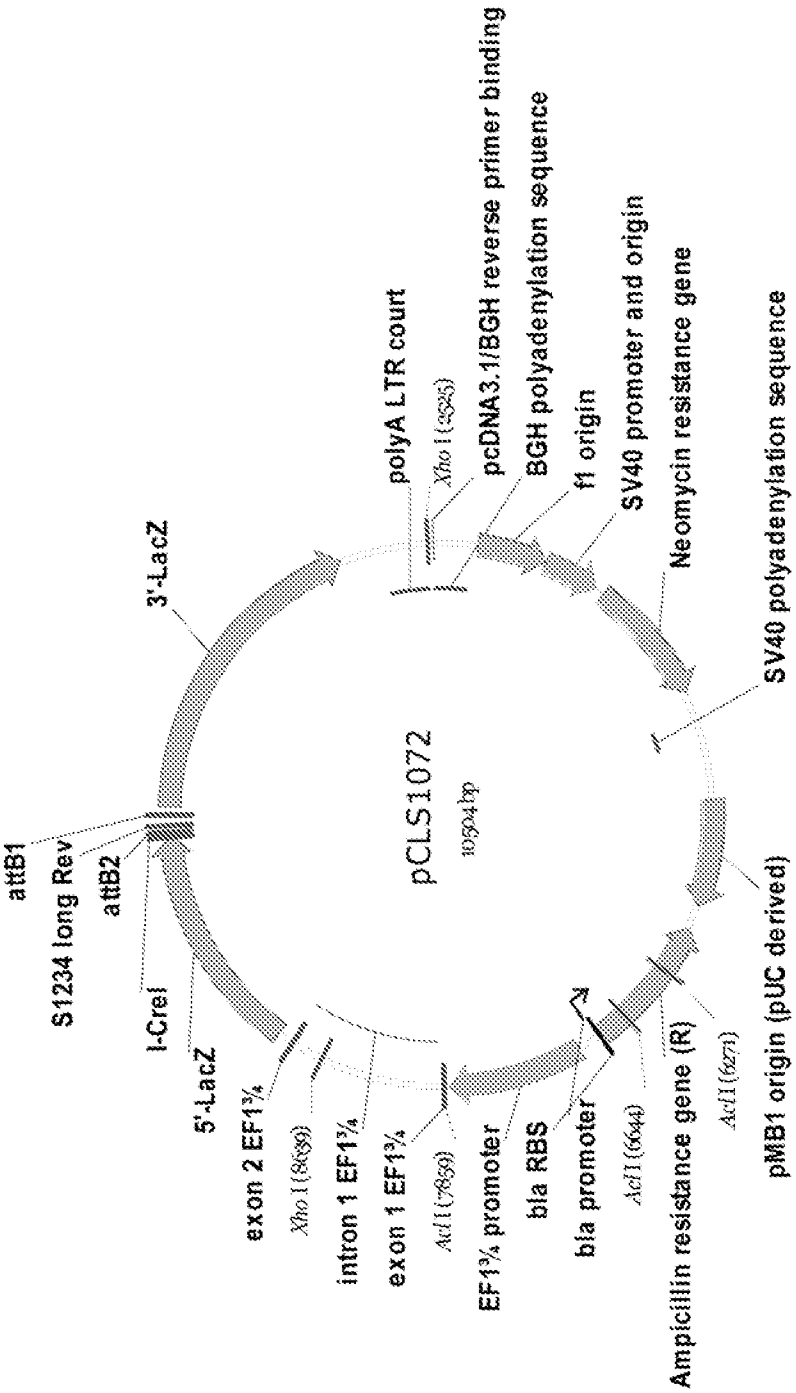
pCLS1072
10504 bp
attB1
S1234 long Rev
attB2
I-CreI
5'-LacZ
exon 2 EF1¾
Xho I (8839)
intron 1 EF1¾
exon 1 EF1¾
Acl I (7859)
EF1¾ promoter
bla RBS
bla promoter
Acl I (6644)
Ampicillin resistance gene (R)
Acl I (6271)
pMB1 origin (pUC derived)
SV40 polyadenylation sequence
Neomycin resistance gene
SV40 promoter and origin
f1 origin
BGH polyadenylation sequence
pcDNA3.1/BGH reverse primer binding
Xho I (8525)
polyA LTR court
3'-LacZ

Figure 8

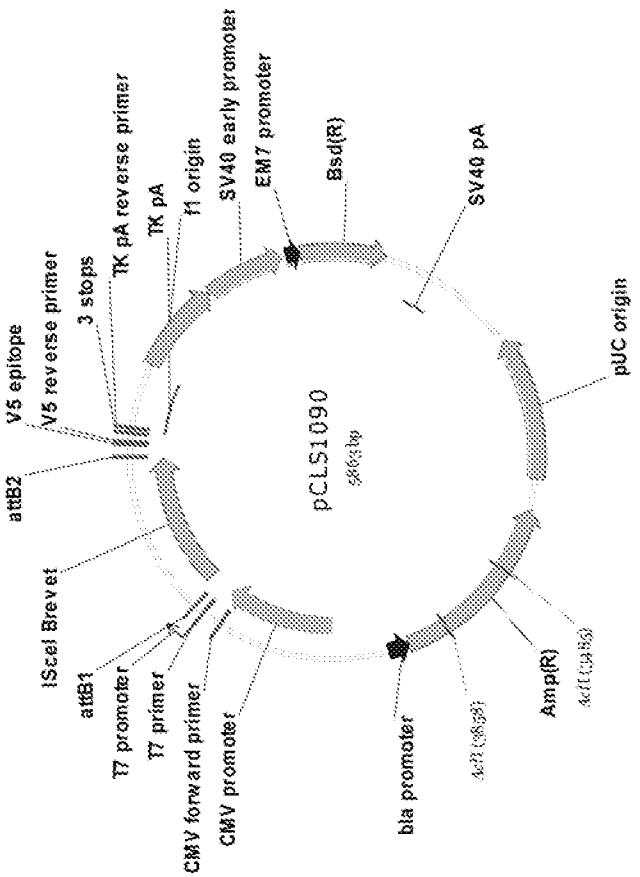
pCLS1090
V5 epitope
V5 reverse primer
3 stops
TK pA reverse primer
TK pA
f1 origin
SV40 early promoter
EM7 promoter
Bsd(R)
SV40 pA
pUC origin
attB2
ISceI BreveI
attB1
T7 promoter
T7 primer
CMV forward primer
CMV promoter
bla promoter
Amp(R)

Figure 9

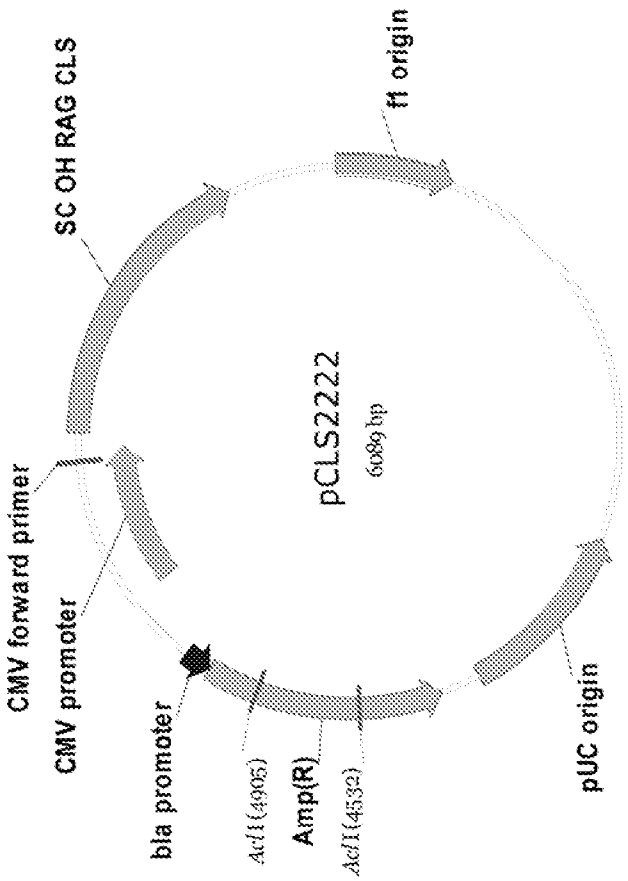
SC OH RAG CLS
f1 origin
CMV forward primer
CMV promoter
pCLS2222
6089 bp
bla promoter
AclI(4905)
Amp(R)
AclI(4532)
pUC origin

Figure 10

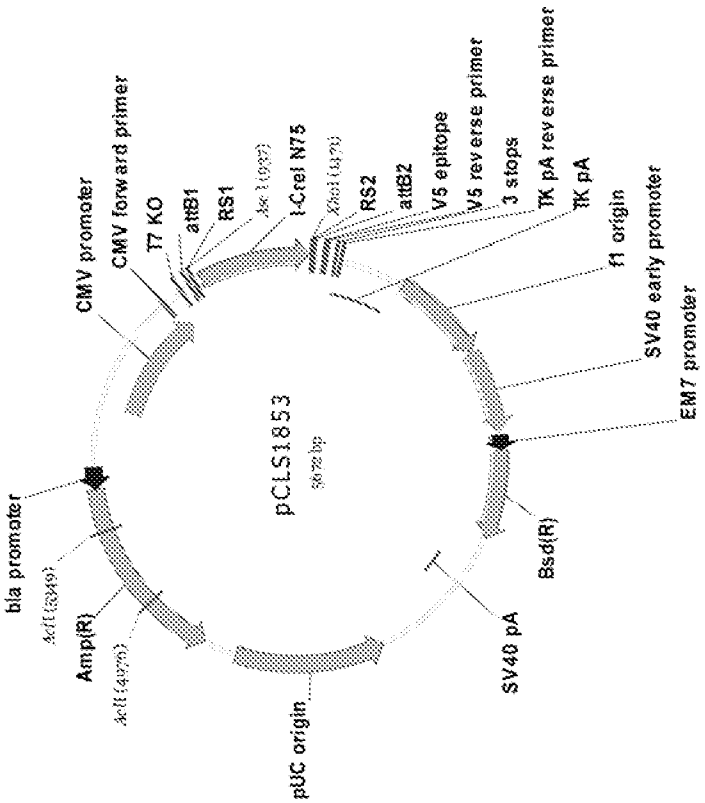
pCLS1853
CMV promoter
CMV forward primer
T7 KO
attB1
RS1
I-CreI N75
RS2
attB2
V5 epitope
V5 reverse primer
3 stops
TK pA reverse primer
TK pA
f1 origin
SV40 early promoter
EM7 promoter
Bsd(R)
SV40 pA
pUC origin
Amp(R)
bla promoter

Figure 11

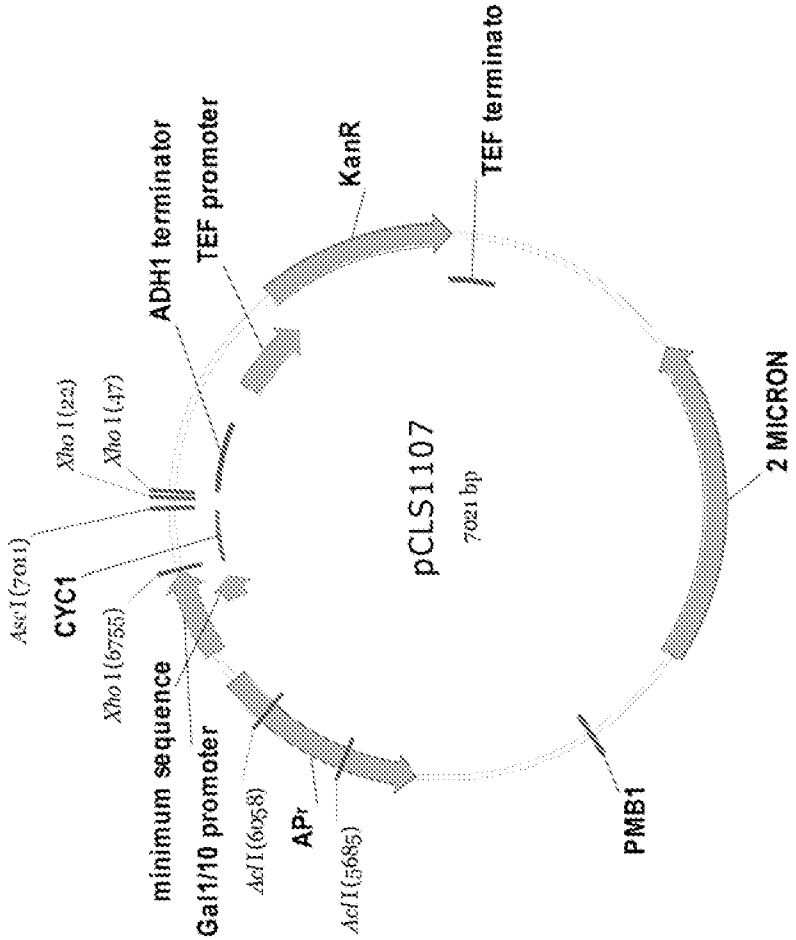
pCLS1107
7021 bp
ADH1 terminator
TEF promoter
KanR
TEF terminato
2 MICRON
PMB1
APr
CYC1
minimum sequence
Gal1/10 promoter
AscI(7011)
XhoI(22)
XhoI(47)
XhoI(6755)
AclI(6958)
AclI(5685)

Figure 12

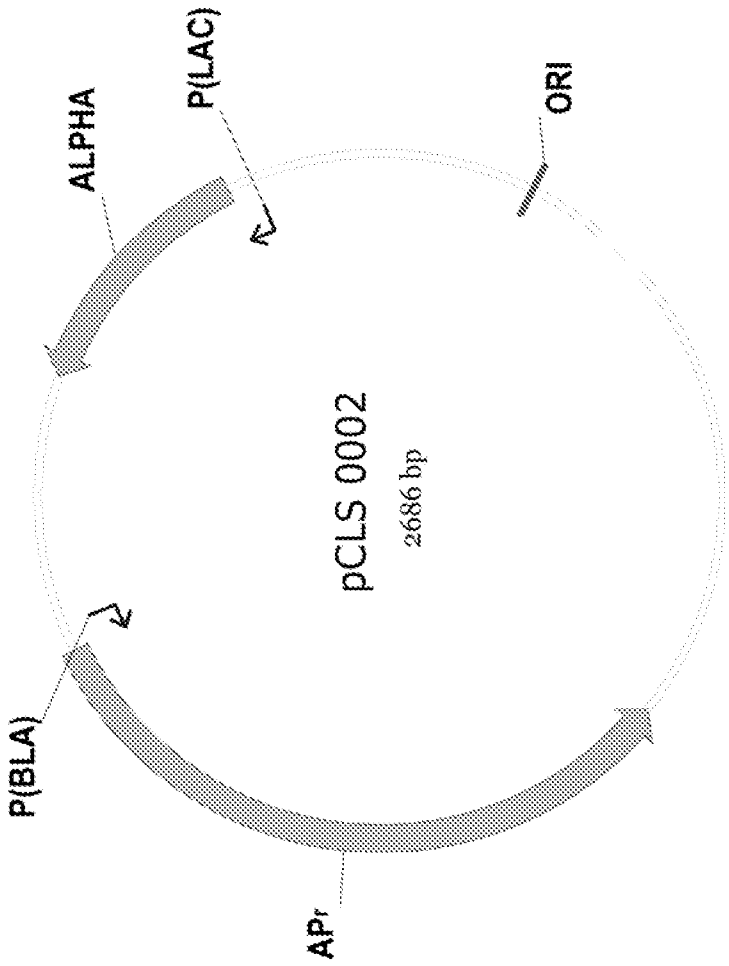


Figure 13

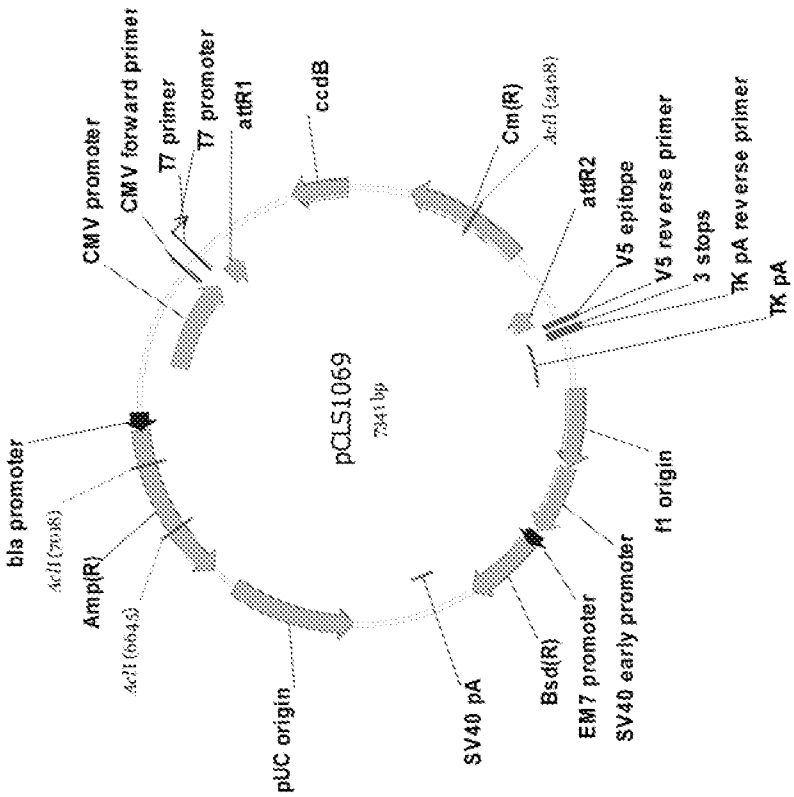
pCLS1069
7341 bp
CMV promoter
CMV forward primer
T7 primer
T7 promoter
attR1
ccdB
Cm(R)
AcI (2468)
attR2
V5 epitope
V5 reverse primer
3 stops
TK pA reverse primer
TK pA
f1 origin
SV40 early promoter
EM7 promoter
Bsd(R)
SV40 pA
pUC origin
AcI (6643)
Amp(R)
AcI (7038)
bla promoter

Figure 14

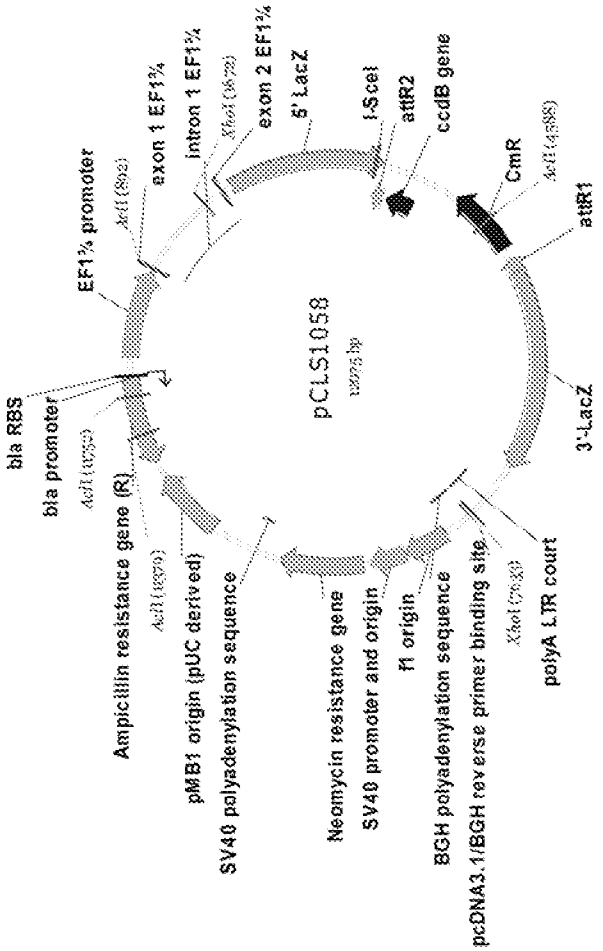
pCLS1058
12975 bp
EF1α promoter
exon 1 EF1α
intron 1 EF1α
exon 2 EF1α
5' LacZ
I-SceI
attR2
ccdB gene
CmR
attR1
3'-LacZ
polyA LTR court
pcDNA3.1/BGH reverse primer binding site
BGH polyadenylation sequence
f1 origin
SV40 promoter and origin
Neomycin resistance gene
SV40 polyadenylation sequence
pMB1 origin (pUC derived)
Ampicillin resistance gene (R)
bla promoter
bla RBS

Figure 15

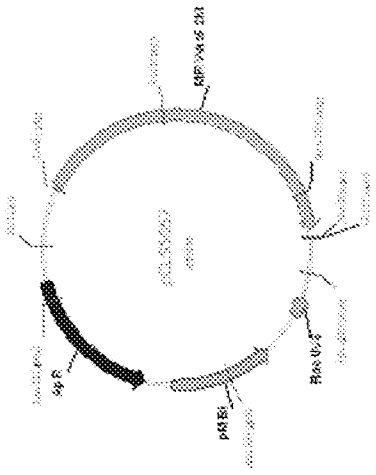

Figure 17

Figure 16

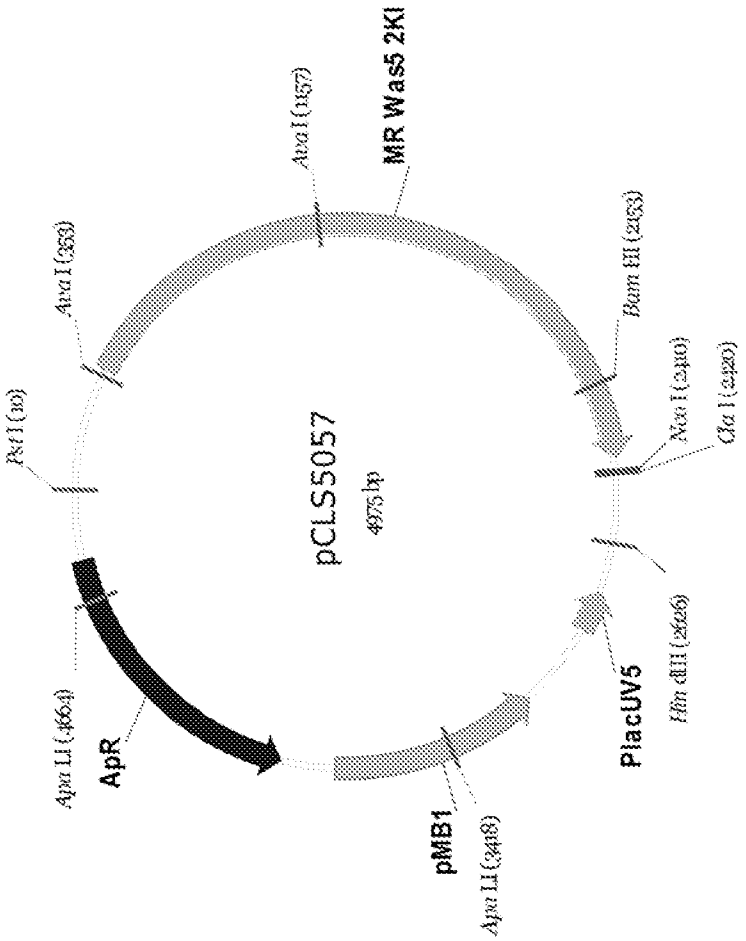
pCLS5057
4975 bp
MR Was5 2KI
PlacUV5
ApR
pMB1
Ava I (957)
Ava I (953)
Pst I (10)
Apa LI (4664)
Apa LI (3448)
Hin dIII (2626)
Nco I (2410)
Cla I (2320)
Bam HI (2453)

Figure 17

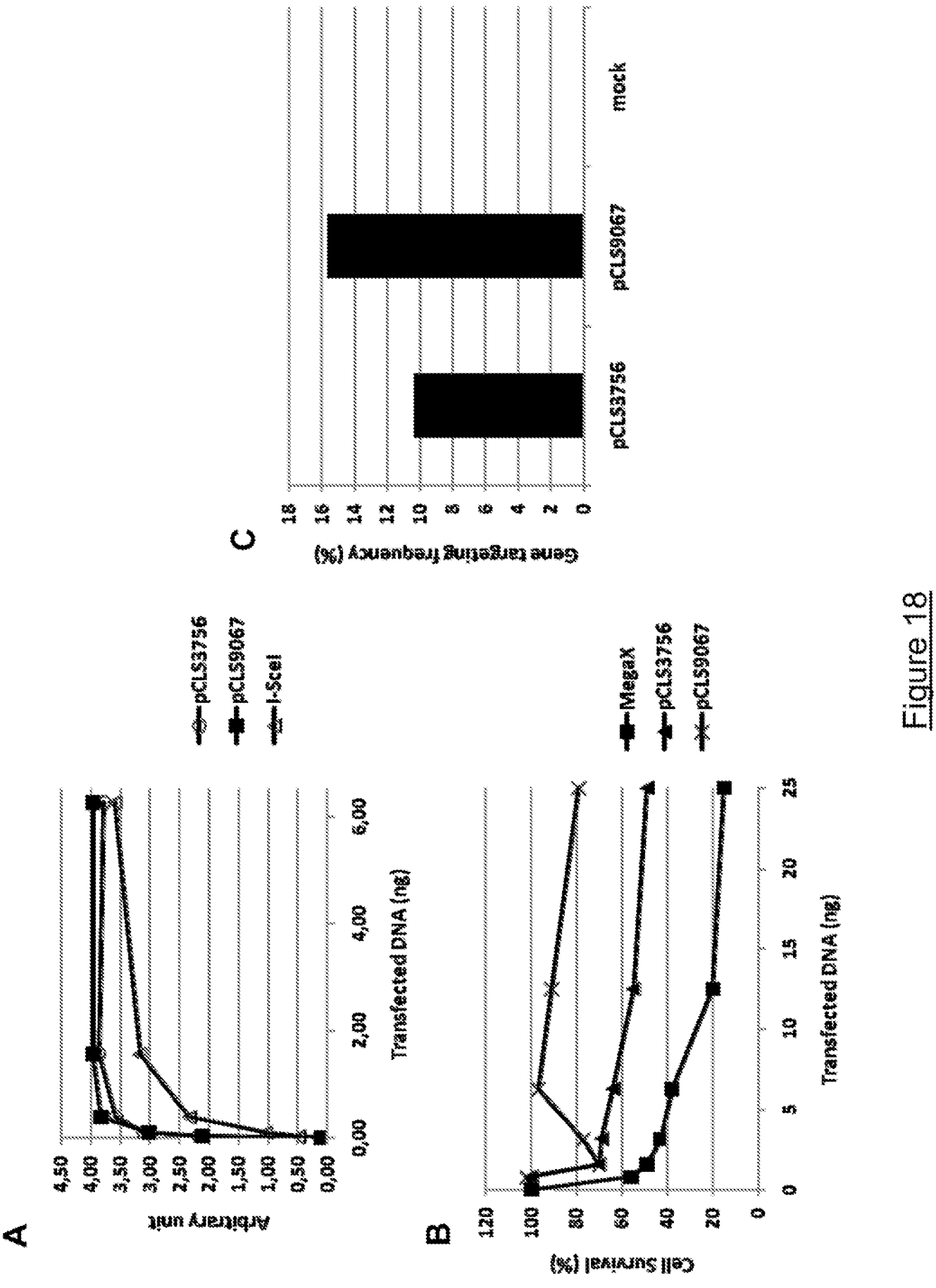
A
Arbitrary unit
4,50
4,00
3,50
3,00
2,50
2,00
1,50
1,00
0,50
0,00
0,00
2,00
4,00
6,00
Transfected DNA (ng)
pCLS3756
pCLS9067
I-SceI
B
Cell Survival (%)
120
100
80
60
40
20
0
0
5
10
15
20
25
Transfected DNA (ng)
MegaX
pCLS3756
pCLS9067
C
Gene targeting frequency (%)
18
16
14
12
10
8
6
4
2
0
pCLS3756
pCLS9067
mock

Figure 18