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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 August 2009 (06.08.2009)

PCT

(10) International Publication Number
WO 2009/095742 A1

(51) International Patent Classification:
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/IB2008/001331

(22) International Filing Date: 31 January 2008 (31.01.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **CELLECTIS** [FR/FR]; 102 route de Noisy, F-93235 CEDEX Romainville (FR).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **GRIZOT, Sylvestre** [FR/FR]; 35 rue de l'Aigle, F-92250 La Garenne Colombes (FR).

(74) Agent: **CABINET ORES**; 36 rue de St Pétersbourg, 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, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(54) **Title:** NEW I-CREI DERIVED SINGLE-CHAIN MEGANUCLEASE AND USES THEREOF

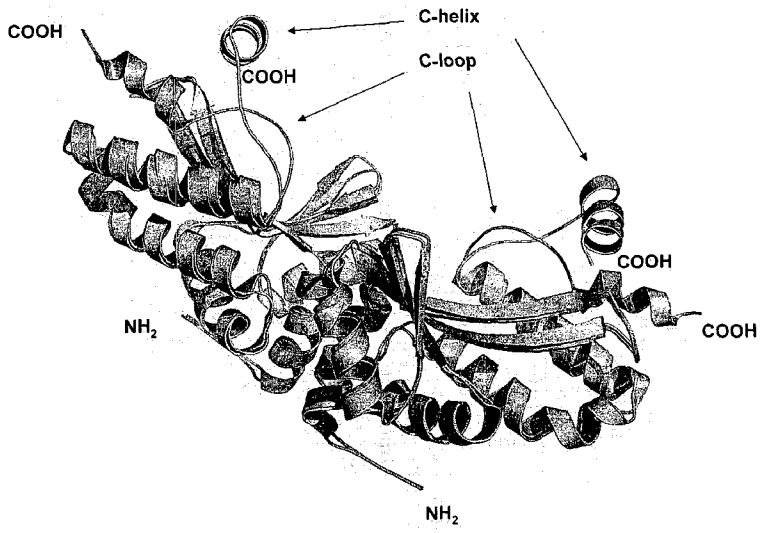

FIGURE 1

(57) **Abstract:** A new I-*Cre*I derived single-chain meganuclease comprising two domains, each domain comprising a portion of a parent I-*Cre*I monomer which extends at least from the beginning of the first alpha helix to the end of the C-terminal loop and said two domains being joined by a peptidic linker which allows them to fold as a I-*Cre*I dimer that is able to bind and cleave a chimeric DNA target comprising one different half of each parent homodimeric I-*Cre*I meganuclease target sequence. Use of said I-*Cre*I derived single-chain meganuclease for genetic engineering, genome therapy and antiviral therapy.

WO 2009/095742 A1

NEW I-*Cre*I DERIVED SINGLE-CHAIN MEGANUCLEASE AND USES THEREOF

The invention relates to a new I-*Cre*I derived single-chain meganuclease, to a vector encoding said new I-*Cre*I derived single-chain meganuclease, to a cell, an animal or a plant modified by said vector and to the use of said I-*Cre*I derived single-chain meganuclease and derived products for genetic engineering, genome therapy and antiviral therapy.

Among the strategies to engineer a given genetic locus, the use of rare cutting DNA endonucleases such as meganucleases has emerged as a powerful tool to increase homologous gene targeting through the generation of a DNA double strand break (DSB). Meganucleases recognize large (>12 bp) sequences, and can therefore cleave their cognate site without affecting global genome integrity. Homing endonucleases, the natural meganucleases, constitute several large families of proteins encoded by mobile introns or inteins. Their target sequence is usually found in homologous alleles that lack the intron or intein, and cleavage initiates the transfer of the mobile element into the broken sequence by a mechanism of DSB-induced homologous recombination. I-*Sce*I was the first homing endonuclease used to stimulate homologous recombination over 1000-fold at a genomic target in mammalian cells (Choulika *et al.*, Mol. Cell. Biol., 1995, 15, 1968-1973; Cohen-Tannoudji *et al.*, Mol. Cell. Biol., 1998, 18, 1444-1448; Donoho *et al.*, Mol. Cell. Biol., 1998, 18, 4070-4078; Alwin *et al.*, Mol. Ther., 2005, 12, 610-617; Porteus, M. H., Mol. Ther., 2006, 13, 438-446; Rouet *et al.*, Mol. Cell. Biol., 1994, 14, 8096-8106). Recently, I-*Sce*I was also used to stimulate targeted recombination in mouse liver *in vivo*, and recombination could be observed in up to 1 % of hepatocytes (Gouble *et al.*, J. Gene Med., 2006, 8, 616-22).

However an inherent limitation of such a methodology is that it requires the prior introduction of the natural cleavage site into the locus of interest. To circumvent this limitation, significant efforts have been made over the past years to generate endonucleases with tailored cleavage specificities. Such proteins could be used to cleave genuine chromosomal sequences and open new perspectives for genome engineering in wide range of applications. For example, meganucleases could be used to induce the correction of mutations linked with monogenic inherited

diseases, and bypass the risk due to the randomly inserted transgenes used in current gene therapy approaches (Hacein-Bey-Abina *et al.*, Science, 2003, 302, 415-419).

Fusion of Zinc-Finger Proteins (ZFPs) with the catalytic domain of the *Fok*I, a class IIS restriction endonuclease, were used 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, 1169-1175 ; 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 nucleases could recently be used for the engineering of the ILR2G gene in human cells from the lymphoid lineage (Urnov *et al.*, Nature, 2005, 435, 646-651).

The binding specificity of Cys2-His2 type Zinc-Finger Proteins, is easy to manipulate, probably because they represent a simple (specificity driven by essentially four residues per finger), and modular system (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 (Rebar, E.J. and C.O. Pabo, Science, 1994, 263, 671-673 ; Kim, J.S. and C.O. 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, 11163-11167 ; 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, 11163-11167 ; Isalan M. and A. Klug, Nat. Biotechnol., 2001, 19, 656-660) laboratories resulted in a large repertoire of novel artificial ZFPs, able to bind most G/ANNG/ANNG/ANN sequences.

Nevertheless, ZFPs might have their limitations, especially for applications requiring a very high level of specificity, such as therapeutic applications. The *Fok*I nuclease activity in fusion acts as a dimer, but it was recently shown that it could cleave DNA when only one out of the two monomer was bound to DNA, or when the two monomers were bound to two distant DNA sequences (Catto *et al.*, Nucleic Acids Res., 2006, 34, 1711-1720). Thus, specificity might be very degenerate, as illustrated by toxicity in mammalian cells (Porteus, M.H. and D. Baltimore, Science, 2003, 300, 763) and Drosophila (Bibikova *et al.*, Genetics, 2002, 161, 1169-1175; Bibikova *et al.*, Science, 2003, 300, 764-.).

Given their exquisite specificity, homing endonucleases may represent ideal scaffolds for engineering tailored endonucleases. Several studies have shown that the DNA binding domain from LAGLIDADG proteins, the most widespread homing endonucleases (Chevalier, B. S. and Stoddard B. L., Nucleic Acids Res. 2001; 29:3757-74) could be engineered. LAGLIDADG refers to the only sequence actually conserved throughout the family and is found in one or more often two copies in the protein (Lucas *et al.*, Nucleic Acids Res., 2001, 29:960-969). Proteins with a single motif, such as I-*Cre*I (Protein Data Bank accession number 1G9Y) and I-*Mso*I, form homodimers and cleave palindromic or pseudo-palindromic DNA sequences, whereas the larger, double motif proteins, such as I-*Sce*I and I-*Dmo*I are monomers and cleave non-palindromic targets. Several different LAGLIDADG proteins have been crystallized, and they exhibit a very striking conservation of the core structure that contrasts with the lack of similarity at the primary sequence level (Jurica *et al.*, Mol. Cell., 1998; 2:469-476; Chevalier *et al.*, Nat. Struct. Biol. 2001; 8:312-316; Chevalier *et al.*, J. Mol. Biol., 2003, 329:253-69, Moure *et al.*, J. Mol. Biol., 2003, 334:685-695; Moure *et al.*, J. Mol. Biol., 2003, 334, 685-695; Moure *et al.*, Nat. Struct. Biol., 2002, 9:764-770; Ichiyanagi *et al.*, J. Mol. Biol., 2000, 300:889-901; Duan *et al.*, Cell, 1997, 89:555-564; Bolduc *et al.*, Genes Dev., 2003, 17:2875-2888; Silva *et al.*, J. Mol. Biol., 1999, 286:1123-1136). In this core structure, two characteristic αββαββα folds, contributed by two monomers, or by two domains in double LAGLIDAG proteins, are facing each other with a two-fold symmetry. DNA binding depends on the four β-strands from each domain, folded into an antiparallel β-sheet, and forming a saddle on the DNA helix major groove. The catalytic core is central, with a contribution of both symmetric monomers/domains. In addition to this core structure, other domains can be found: for example, PI-*Sce*I, an intein, has a protein splicing domain, and an additional DNA-binding domain (Moure *et al.*, Nat. Struct. Biol., 2002, 9:764-70, Grindl *et al.*, Nucleic Acids Res. 1998, 26:1857-1862). Each I-*Cre*I monomer comprises a C-terminal subdomain made of three helices (alpha 4 (α_4), alpha 5(α_5) and alpha 6(α_6)) and a loop between the α_5 and α_6 helices (C-terminal loop; Figure 1), which is essential for site recognition, DNA binding and cleavage (Prieto *et al.*, Nucleic Acids Res., 2007,, 35, 3267-3271).

Several LAGLIDAG proteins, including PI-*Sce*I (Gimble *et al.*, J. Mol. Biol., 2003, 334:993-1008), I-*Cre*I (Seligman *et al.*, Nucleic Acids Res. 2002, 30:3870-3879; Sussman *et al.*, J. Mol. Biol., 2004, 342:31-41; International PCT Applications WO 2006/097784, WO 2006/097853, WO 2007/060495 and WO 2007/049156; 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, e149), I-*Sce*I (Doyon *et al.*, J Am Chem Soc., 2006, 128:2477-2484) and I-*Mso*I (Ashworth *et al.*, Nature, 2006, 441:656-659) could be modified by rational or semi-retional mutagenesis and screening to acquire new binding or cleavage specificities.

Another strategy was the creation of new meganucleases by domain swapping between I-*Cre*I and I-*Dmo*I, leading to the generation of a meganuclease cleaving the hybrid sequence corresponding to the fusion of the two half parent target sequences (Epinat *et al.*, Nucleic Acids Res., 2003, 31:2952-2962; Chevalier *et al.*, Mol. Cell. 2002, 10:895-905; International PCT Applications WO 03/078619 and WO 2004/031346).

Recently, semi rational design assisted by high throughput screening methods allowed to derive thousands of novel proteins from I-*Cre*I (Smith *et al.*, Nucleic Acids Res. 2006, 34, e149; Arnould *et al.*, J. Mol. Biol., 2006, 355:443-458; International PCT Applications WO 2006/097784, WO 2006/097853, WO 2007/060495 and WO 2007/049156). In such an approach, a limited set of protein residues are chosen after examination of protein/DNA cocrystal structure, and randomized. Coupled with high-throughput screening (HTS) techniques, this method can rapidly result in the identification of hundreds of homing endonucleases derivatives with modified specificities.

Furthermore, DNA-binding sub-domains that were independent enough to allow for a combinatorial assembly of mutations were identified (Smith *et al.*, Nucleic Acids Res. 2006, 34, e149; International PCT Applications WO 2007/049095 and WO 2007/057781). These findings allowed for the production of a second generation of engineered I-*Cre*I derivatives, cleaving chosen targets. This combinatorial strategy, has been illustrated by the generation of meganucleases cleaving a natural DNA target sequence located within the human RAG1 and XPC genes (Smith *et al.*, Nucleic Acids Res., 2006, 34, e149; Arnould *et al.*, J. Mol. Biol.,

2007, 371:49-65; International PCT Applications WO 2007/093836, WO 2007/093918 and WO 2008/010093).

The engineered meganucleases cleaving natural DNA targets from the human RAG1 or XPC genes are heterodimers, which include two separately engineered monomers, each binding one half of the target. Heterodimer formation is obtained by co-expression of the two monomers in the same cells (Smith *et al.*, Nucleic Acids Res., 2006, 34, e149). Such co-expression of two monomers result in the formation of three molecular species: the heterodimer, which cleaves the target of interest, and the two homodimers, which can be considered as by-products (Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458; International PCT Applications WO 2006/097854 and WO 2007/057781; Smith *et al.*, Nucleic Acids Res., 2006, 34, e149). However, individual homodimers can sometimes be responsible for high levels of toxicity (Bibikova *et al.*, Genetics, 2002, 161, 1169-1175; Beumer *et al.*, Genetics, 2006, 172, 2391-2403). Thus, a limiting factor that still remains for the widespread use of the single-LAGLIDADG homing endonucleases such as I-*Cre*I, is the fact that proteins are homodimers. Two possibilities can overcome this issue: either the making of obligatory heterodimers, by redesign of the dimerization interface, resulting in the suppression of functional homodimer formation, or the fusion of the two monomers with a protein link or linker in a single chain molecule, in order to favor intramolecular over intermolecular interactions of distincts αββαββα folds. This last strategy would have an additional advantage in terms of vectorization, for only one coding sequence or proteins would have to be introduced into the nucleus. It would thus alleviate the use of two different vectors, or of bicistronic ones. Note that at least in theory, the single chain molecule would not necessarily alleviate the interaction of αββαββα folds from distinct molecules, especially if the linker is long and flexible, but it should at least favor interactions between αββαββα folds from a same molecule. Furthermore, the making of single chain molecule and the redesign of the dimerization interface are not exclusive strategies and could be used jointly.

The making of a single-chain version of I-*Cre*I (scI-*Cre*I) has already been tested (Epinat *et al.*, Nucleic Acids Res., 2003, 3, 2952-2962, International PCT Application WO 03/078619). In this first version, the N-terminal domain of the single-chain meganuclease (positions 1 to 93 of I-*Cre*I amino acid sequence) consisted

essentially of the αββαββα fold (core domain) of an I-*Cre*I monomer whereas the C-terminal (positions 8 to 163 of I-*Cre*I amino acid sequence) was a nearly complete I-*Cre*I monomer. The linker (MLERIRLFNMR; SEQ ID NO: 1) was derived from the loop joining the two domains of I-*Dmo*I.

Although the first scI-*Cre*I was a functional meganuclease, it was less stable than I-*Cre*I, probably due a less optimal folding as compared to its natural counterpart. The design of the first scI-*Cre*I matched the structure of double LAGLIDADG endonucleases and therefore differed from that of I-*Cre*I in that the N-terminal domain is shorter than the C-terminal domain and lacks the C-terminal subdomain made of three α-helices which may be present in the C-terminal domain of some double LAGLIDADG endonucleases.

The removal of the three C-terminal helices from the first monomer might affect the folding, stability and consequently the cleavage activity of the first scI-*Cre*I since recent works have shown the crucial role of the C-terminal subdomain of I-*Cre*I in the protein DNA binding properties and DNA target cleavage activity (Prieto *et al.*, Nucleic Acids Res., 2007,), 35, 3267-3271).

However, the three C-terminal helices terminate at opposite sides of the dimer structure of I-*Cre*I, far away from the N-terminal helices comprising the LAGLIDADG motif (Figure 1). The length of a flexible linker connecting the C-terminal residue of one domain to the N-terminal residue of the other domain (end to end fusion) would be considerable. Besides engineering such linker would be difficult due to the necessity to go across a large part of the protein surface. Therefore, it is uncertain that proper packing be obtained.

Here the inventors present a new way to design a single chain molecule derived from the I-*Cre*I homodimeric meganuclease. This strategy preserves the core αββαββα (also named as $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$) fold as well as the C-terminal part of the two linked I-*Cre*I units.

The subject matter of the present invention is a single-chain I-*Cre*I meganuclease (scI-CreI) comprising two domains (N-terminal and C-terminal) joined by a peptidic linker, wherein:

(a) each domain, derived from a parent I-*Cre*I monomer, comprises a portion of said parent I-*Cre*I monomer which extends at least from the beginning of

the first alpha helix (α_1) to the end of the C-terminal loop of I-*Cre*I and includes successively: the $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ core domain, the α_4 and α_5 helices and the C-terminal loop of I-*Cre*I, and

(b) the two domains are joined by a peptidic linker which allows said two domains to fold as a I-*Cre*I dimer that is able to bind and cleave a chimeric DNA target comprising one different half of each parent homodimeric I-*Cre*I meganuclease target sequence.

Definitions

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

- 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 "meganuclease" is intended an endonuclease having a double-stranded DNA target sequence of 12 to 45 bp. Said meganuclease is either a dimeric enzyme, wherein each domain is on a monomer or a monomeric enzyme comprising the two domains on a single polypeptide. Said meganuclease may be derived from a LAGLIDADG homing endonuclease, for example from I-*Cre*I.

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

sequences. The single-chain meganuclease is also named single-chain derivative, single-chain meganuclease, single-chain meganuclease derivative or chimeric meganuclease.

- by "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 core domain comprises four beta-strands ($\beta_1\beta_2\beta_3\beta_4$) folded in an antiparallel beta-sheet which interacts with one half of the DNA target. This core domain is able to associate with another LAGLIDADG homing endonuclease core domain which 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-*Cre*I (163 amino acids), the core domain comprises the residues 6 to 94 of I-*Cre*I.

- 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 "subdomain" 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 "meganuclease variant" or "variant" is intented a meganuclease obtained by replacement of at least one residue in the amino acid sequence of the wild-type meganuclease (natural meganuclease) with a different amino acid.

- 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 "meganuclease 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.

- by "I-*Cre*I" is intended the wild-type I-*Cre*I (Protein Data Bank accession number 1g9y), corresponding to the sequence SEQ ID NO: 20 in the sequence listing.

- by "parent I-*Cre*I monomer"or "I-*Cre*I monomer" is intended one domain of the homodimeric wild-type I-*Cre*I meganuclease or of a functional (homodimeric) variant thereof, corresponding to the full-length I-*Cre*I amino acid sequence (positions 1 to 163 of SEQ ID NO: 20)

- by "portion of said parent I-*Cre*I monomer which extends at least from the beginning of the first alpha helix to the end of the C-terminal loop of I-*Cre*I and includes successively: the $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ core domain, the α_4 and α_5 helices and the C-terminal loop" is intended the amino acid sequence corresponding to at least positions 8 to 143 of I-*Cre*I.

- by "I-*Cre*I site" is intended a 22 to 24 bp double-stranded DNA sequence which is cleaved by I-*Cre*I. I-*Cre*I sites include the wild-type (natural) non-palindromic I-*Cre*I homing site and the derived palindromic sequences such as the sequence 5'- $c_{-11}a_{-10}a_{-9}a_{-8}a_{-7}c_{-6}g_{-5}t_{-4}c_{-3}g_{-2}t_{-1}a_{+1}c_{+2}g_{+3}a_{+4}c_{+5}g_{+6}t_{+7}t_{+8}t_{+9}t_{+10}g_{+11}$ also called C1221 (SEQ ID NO :21; figure 3).

- 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-*Cre*I, or a variant, or a single-chain chimeric meganuclease derived from I-*Cre*I. 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 indicated above for C1221. Cleavage of the DNA target occurs at the nucleotide positions +2 and -2, respectively for the sense and the antisense strand. Unless otherwiwe indicated, the position at which cleavage of the DNA target by an I-*Cre* I meganuclease variant or a

single-chain derivative 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 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 "chimeric DNA target comprising one different half of each parent homodimeric I-*Cre*I meganuclease target sequence" is intended the target sequence comprising the left part of the palindromic target sequence cleaved by the homodimeric meganuclease made of two identical monomers of one parent monomer and the right part of the palindromic target sequence cleaved by the homodimeric meganuclease made of two identical monomers of the other parent monomer.

- 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 a homologous recombination between sequences, more particularly having at least 95 % identity, preferably 97 % identity and more preferably 99 %.

- "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 settings.

- "individual" includes mammals, as well as other vertebrates (e.g., birds, fish and reptiles). The terms "mammal" and "mammalian", as used herein, refer

to any vertebrate animal, including monotremes, marsupials and placental, that suckle their young and either give birth to living young (eutharian or placental mammals) or are egg-laying (metatharian or nonplacental mammals). Examples of mammalian species include humans and other primates (e.g., monkeys, chimpanzees), rodents (e.g., rats, mice, guinea pigs) and others such as for example: cows, pigs and horses.

- by "mutation" is intended the substitution, deletion, insertion of one 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.

- by "site-specific mutation" is intended the mutation of a specific nucleotide/codon in a nucleotidic sequence as opposed to random mutation.

The single-chain I-*Cre*I meganuclease according to the invention is also named, scI-*Cre*I meganuclease or sc-I-*Cre*I.

According to the present invention, the sequence of the linker is chosen so as to allow the two domains of the sc-I-*Cre*I to fold as a I-*Cre*I dimer and to bind and cleave a chimeric DNA target comprising one different half of each parent homodimeric I-*Cre*I meganuclease target sequence.

I-*Cre*I dimer formation may be assessed by well-known assays such as sedimentation equilibrium experiments performed by analytical centrifugation, as previously described in Prieto *et al.*, Nucleic Acids Research, 2007, 35, 3267-3271.

DNA binding may be assessed by well-known assays such as for example, electrophoretic mobility shift assays (EMSA), as previously described in Prieto *et al.*, Nucleic Acids Research, 2007, 35, 3267-3271.

The cleavage activity of the single-chain meganuclease according to the invention may be measured 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; Arnould *et al.*, J. Mol. Biol., Epub 10 May 2007.

For example, the cleavage activity of the single-chain meganuclease according to the present invention may be measured by a direct repeat recombination

assay, in yeast or mammalian cells, using a reporter vector, by comparison with that of the corresponding heterodimeric meganuclease or of another single-chain meganuclease, derived from identical parent I-*Cre*I monomers. The reporter vector comprises two truncated, non-functional copies of a reporter gene (direct repeats) and the genomic (non-palindromic) DNA target sequence -comprising one different half of each (palindromic or pseudo-palindromic) parent homodimeric I-*Cre*I meganuclease target sequence, within the intervening sequence, cloned in a yeast or a mammalian expression vector. Expression of the meganuclease results in cleavage of the genomic chimeric DNA target sequence. This cleavage induces homologous recombination between the direct repeats, resulting in a functional reporter gene (LacZ, for example), whose expression can be monitored by appropriate assay. In addition, the activity of the single-chain meganuclease towards its genomic DNA target can be compared to that of I-*Cre*I towards the I-*Cre*I site, at the same genomic locus, using a chromosomal assay in mammalian cells (Arnould *et al.*, J. Mol. Biol., Epub 10 May 2007). In addition the specificity of the cleavage by the single-chain meganuclease may be assessed by comparing the cleavage of the chimeric DNA target sequence with that of the two palindromic sequences cleaved by the parent I-*Cre*I homodimeric meganucleases.

The N-terminal sequence of the two domains of the sc-I-*Cre*I may start at position 1, 2, 3, 4, 5, 6 or 8 of I-*CreI*. In a preferred embodiment, the N-terminal domain starts at position 1 or 6 of I-*Cre*I and the C-terminal domain starts at position 2 or 6 of I-*Cre*I.

The C-terminal sequence of the two domains terminates just after the C-terminal loop, for example at position 143 or 145 of I-*Cre*I. Alternatively, the C-terminal sequence of the domain(s) further includes at least the alpha 6 helix (positions 145 to 150 of I-*Cre*I) and eventually additional C-terminal residues. In this case, the C-terminal sequence of the domain(s) may terminates at position 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162 or 163 of I-*Cre*I. Preferably at position 152 to 160. For example at position 152, 156, 160 or 163. In a preferred embodiment, the N-terminal domain terminates at position 163 of I-*Cre*I. More preferably, both domains terminate at positions 163 of I-*Cre*I.

In another preferred embodiment of the sc-I-*Cre*I meganuclease, the linker comprises or consists of a linear sequence of 10 to 100 consecutive amino acids, preferably 15 to 50 amino acids, more preferably 20 to 35 amino acids. The linker is advantageously selected from the group consisting of the sequences comprising or consisting of SEQ ID NO: 2 to 19. Preferably, it is selected from the group consisting of the sequences of SEQ ID NO: 2 to 12 and 14 to 19. more preferably, the linker consists of the sequence SEQ ID NO: 2.

The single-chain I-*Cre*I meganuclease according to the invention may be derived from wild-type I-*Cre*I monomers or functional variants thereof. In addition, one or more residues inserted at the NH_2 terminus and/or COOH terminus of the parent monomers. Additional codons may be added at the 5' or 3' end of the parent monomers to introduce restrictions sites which are used for cloning into various vectors. An example of said sequence is SEQ ID NO: 23 which has an alanine (A) residue inserted after the first methionine residue and an alanine and an aspartic acid (AD) residues inserted after the C-terminal proline residue. These sequences allow having DNA coding sequences comprising the *Nco*I (ccatgg) and *Eag*I (cggccg) restriction sites which are used for cloning into various vectors. For example, a tag (epitope or polyhistidine sequence) may also be introduced at the NH_2 terminus of the N-terminal domain and/or COOH terminus of C-terminal domain; said tag is useful for the detection and/or the purification of said sc-I-*Cre*I meganuclease.

Preferably, the sc-I-*Cre*I meganucleaseis derived from the monomers of an heterodimeric I-*Cre*I variant, more preferably of a variant having novel cleavage specificity as previously described (Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458; Smith *et al.*, Nucleic Acids Res., 2006, 34, e149 ; Arnould *et al.*, J. Mol. Biol., Epub 10 May 2007; International PCT Applications WO 2006/097784, WO 2006/097853, WO 2007/049095, WO 2007/057781, WO 2007/060495, WO 2007/049156, WO 2007/093836 and WO 2007/093918)

Therefore the domains of the single-chain I-*Cre*I meganuclease may comprise mutations at positions of I-*Cre*I 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).

Preferably said mutations modify the cleavage specificity of the meganuclease and result in a meganuclease with novel specificity, which is able to cleave a DNA target from a gene of interest. More preferably, said mutations are substitutions of one or more amino acids in a first functional subdomain corresponding to that situated from positions 26 to 40 of I-*Cre*I amino acid sequence, that alter the specificity towards the nucleotide at positions ± 8 to 10 of the DNA target, and/or substitutions in a second functional subdomain corresponding to that situated from positions 44 to 77 of I-*Cre*I amino acid sequence, that alter the specificity towards the nucleotide at positions ± 3 to 5 of the DNA target, as described previously (International PCT Applications WO 2006/097784, WO 2006/097853, WO 2007/060495, WO 2007/049156, WO 2007/049095 and WO 2007/057781; Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458; Smith *et al.*, Nucleic Acids Res., 2006). The substitutions correspond advantageously to positions 26, 28, 30, 32, 33, 38, and/or 40, 44, 68, 70, 75 and/or 77 of I-*Cre*I amino acid sequence. For cleaving a DNA target, wherein n_{-4} is t or n_{+4} is a, said I-*Cre*I domain has advantageously a glutamine (Q) in position 44; for cleaving a DNA target, wherein n_{-4} is a or n_{+4} is t, said domain has an alanine (A) or an asparagine in position 44, and for cleaving a DNA target, wherein n_{-9} is g or n_{+9} is c, said domain has advantageously an arginine (R) or a lysine (K) in position 38.

According to another preferred embodiment of said scI-*Cre*I meganuclease, at least one domain has mutations at positions 26 to 40 and/or 44 to 77 of I-*Cre*I, said scI-*Cre*I meganuclease being able to cleave a non-palindromic DNA sequence, wherein at least the nucleotides at positions +3 to +5, +8 to +10, -10 to -8 and -5 to -3 of said DNA sequence correspond to the nucleotides at positions +3 to +5, +8 to +10 , -10 to -8 and -5 to -3 of a DNA target from a gene of interest. Preferably, both domains of the scI-*Cre*I meganuclease are mutated at positions 26 to 40 and/or 44 to 77. More preferably, both domains have different mutations at positions 26 to 40 and 44 to 77 of I-*Cre*I.

In another preferred embodiment of said scI-*Cre*I meganuclease, at least one domain comprises one or more mutations at positions of other amino acid residues of I-*Cre*I that interact with the DNA target sequence. 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., 2007, 35, 3262-3271). 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-*Cre*I. Two residues may be mutated in one domain 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-*Cre*I site. Preferably, the residue at position 138 or 139 is substituted by an 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 modifies the specificity of the scI-*Cre*I meganuclease towards the nucleotide at positions ± 1 to 2, ± 6 to 7 and/or ± 11 to 12 of the I-*Cre*I site.

In another preferred embodiment of said scI-*Cre*I meganuclease, at least one domain comprises one or more additional mutations that improve the binding and/or the cleavage properties, including the cleavage activity and/or specificity of the scI-*Cre*I meganuclease towards the DNA target sequence from a gene of interest. The additional residues which are mutated may be on the entire I-*Cre*I sequence.

According to a more preferred embodiment of said sc-I-*Cre*I meganuclease, said additional mutation(s) impair(s) the formation of functional homodimers from the domains of the sc-I-*Cre*I meganuclease.

Each parent monomer has at least two residues Z and Z' of the dimerisation interface which interact with residues Z' and Z, respectively of the same or another parent monomer (two pairs ZZ' of interacting residues) to form two homodimers and one heterodimer. According to the present invention, one of the two pairs of interacting residues of the dimerisation interface is swapped to obtain a first domain having two residues Z or Z' and a second domain having two residues Z' or Z, respectively. As a result, the first and the second domains each having two residues Z or two residues Z' can less easily homodimerize (inter-sc-I-CreI domains interaction)

than their parent counterpart, whereas the presence of two pairs ZZ' of interacting residues at the interface of the two domains of the sc-I-*Cre*I makes the heterodimer formation (intra-sc-I-CreI domains interaction) favourable.

Therefore the domains of the sc-I-CreI meganuclease have advantageously at least one of the following pairs of mutations, respectively for the first (N-terminal or C-terminal domain) and the second domain (C-terminal domain or N-terminal domain) :

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

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

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

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

For example, the first domain may have the mutation D137R and the second domain, the mutation R51D.

Alternatively, the sc-I-CreI meganuclease comprises 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 domain comprises the substitution of the lysine residues at positions 7 and 96 by an acidic amino acid and the other

domain comprises the substitution of the glutamic acid residues at positions 8 and 61 by a basic amino acid. More preferably, the sc-I-CreI meganuclease comprises three pairs of mutations as defined in a), b) and c), above. The sc-I-CreI meganuclease consists advantageously of a first domain (A) having at least the mutations selected from: (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 domain (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.

Another example of mutations that impair the formation of functional homodimers from the domains of the sc-I-*Cre*I meganuclease is the G19S mutation. The G19S mutation is advantageously introduced in one of the two domains of the sc-I-*Cre*I meganuclease, so as to obtain a single-chain meganuclease having enhanced cleavage activity and enhanced cleavage specificity. In addition, to enhance the cleavage specificity further, the other domain may carry a distinct mutation that impairs the formation of a functional homodimer or favors the formation of the heterodimeric sc-I-CreI meganuclease, as defined above.

In another preferred embodiment of said sc-I-*Cre*I meganuclease, said mutations 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 and W.

The subject-matter of the present invention is also a polynucleotide fragment encoding a single-chain meganuclease as defined above. According to a preferred embodiment of said polynucleotide, the nucleic acid sequences encoding the two I-*Cre*I domains of said single-chain meganuclease have less than 80 % nucleic sequence identity, preferably less than 70 % nucleic sequence identity. This reduces the risk of recombination between the two sequences and as a result, the genetic stability of the polynucleotide construct and the derived vector is thus increased. This may be obtained by rewriting the I-*Cre*I sequence using the codon usage and the genetic code degeneracy. For example, one of the domains is derived from the wild-type I-*Cre*I coding sequence (SEQ ID NO: 22) and the other domain is derived from a rewritten version of the I-*Cre*I coding sequence (SEQ ID NO: 24). Furthermore the

codons of the cDNAs encoding the single-chain meganuclease are chosen to favour the expression of said proteins in the desired expression system.

The subject-matter of the present invention is also a recombinant vector for the expression of a single-chain meganuclease according to the invention. The recombinant vector comprises at least one polynucleotide fragment encoding a single-chain meganuclease, as defined above.

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 consists of a 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 of skill 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, 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).

Preferred vectors include lentiviral vectors, and particularly self inactivacting lentiviral vectors.

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 single-chain meganuclease of the invention is placed under control of appropriate transcriptional and translational control elements to permit production or synthesis of said meganuclease. Therefore, said polynucleotide is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a promoter operatively linked to said encoding 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. 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-β-D-thiogalacto-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), α-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.

Alternatively, the vector coding for the 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).

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 pb to 6000 pb, more preferably from 1000 pb to 2000 pb. 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 is preferably a sequence which repairs a mutation in the gene of interest (gene correction or recovery of a functional gene), for the purpose of genome therapy. Alternatively, it can be any other sequence used to alter the chromosomal DNA in some specific way including a sequence used to modify a specific sequence, to attenuate or activate the gene of interest, to inactivate or delete the gene of interest 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/human recombinant cell lines).

The invention also concerns a prokaryotic or eukaryotic host cell which 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 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 eukaryotic cell, such as an animal, plant or yeast cell.

The subject-matter of the present invention is further the use of a meganuclease, a polynucleotide, preferably included in an expression vector, a cell, a transgenic plant, a non-human transgenic mammal, as defined above, for molecular biology, for *in vivo* or *in vitro* genetic engineering, and for *in vivo* or *in vitro* genome engineering, for non-therapeutic purposes.

Molecular biology includes with no limitations, DNA restriction and DNA mapping. Genetic and genome engineering for non therapeutic purposes include for example (i) gene targeting of specific loci in cell packaging lines for protein production, (ii) gene targeting of specific loci in crop plants, for strain improvements and metabolic engineering, (iii) targeted recombination for the removal of markers in genetically modified crop plants, (iv) targeted recombination for the removal of markers in genetically modified microorganism strains (for antibiotic production for example).

According to an advantageous embodiment of said use, it is for inducing a double-strand break in a site of interest comprising a 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, modifying a specific sequence, restoring a functional gene in place of a mutated one, attenuating or activating an endogenous gene of interest, introducing a mutation into a site of interest, introducing an exogenous gene or a part thereof, inactivating or detecting an endogenous gene or a part thereof, translocating a chromosomal arm, or leaving the DNA unrepaired and degraded.

The subject-matter of the present invention is also a method of genetic engineering, characterized in that it comprises a step of double-strand nucleic acid breaking in a site of interest located on a vector comprising a DNA target as defined hereabove, by contacting said vector with a meganuclease as defined above, thereby inducing an homologous recombination with another vector presenting homology with the sequence surrounding the cleavage site of said meganuclease.

The subjet-matter of the present invention is also a method of genome engineering, characterized in that it comprises the following steps: 1) double-strand breaking a genomic locus comprising at least one DNA target of a meganuclease as defined above, by contacting said target with said meganuclease; 2) maintaining said broken genomic locus under conditions appropriate for homologous recombination with a targeting DNA construct comprising the sequence to be introduced in said locus, flanked by sequences sharing homologies with the targeted locus.

The subject-matter of the present invention is also a method of genome engineering, characterized in that it comprises the following steps: 1) double-strand breaking a genomic locus comprising at least one DNA target of a meganuclease as defined above, by contacting said cleavage site with said meganuclease; 2) maintaining said broken genomic locus under conditions appropriate for homologous recombination with chromosomal DNA sharing homologies to regions surrounding the cleavage site.

The subject-matter of the present invention is also the use of at least one meganuclease as defined above, one or two derived polynucleotide(s), preferably

included in expression vector(s), as defined above, for the preparation of a medicament for preventing, improving or curing a genetic disease in an individual in need thereof, said medicament being administrated by any means to said individual.

The subject-matter of the present invention is also a method for preventing, improving or curing a genetic disease in an individual in need thereof, said method comprising the step of administering to said individual a composition comprising at least a meganuclease as defined above, by any means.

In this case, the use of the meganuclease as defined above, comprises at least the step of (a) inducing in somatic tissue(s) of the individual a double stranded cleavage at a site of interest of a gene comprising at least one recognition and cleavage site of said meganuclease, and (b) introducing into the individual 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. The targeting DNA is introduced into the individual under conditions appropriate for introduction of the targeting DNA into the site of interest.

According to the present invention, said double-stranded cleavage is induced, either *in toto* by administration of said meganuclease to an individual, or *ex vivo* by introduction of said meganuclease into somatic cells removed from an individual and returned into the individual after modification.

In a preferred embodiment of said use, the meganuclease is combined with a targeting DNA construct comprising a sequence which repairs a mutation in the gene flanked by sequences sharing homologies with the regions of the gene surrounding the genomic DNA cleavage site of said meganuclease, as defined above. The sequence which repairs the mutation is either a fragment of the gene with the correct sequence or an exon knock-in construct.

For correcting a gene, cleavage of the gene occurs in the vicinity of the mutation, preferably, within 500 bp of the mutation. The targeting construct comprises a gene fragment which has at least 200 bp of homologous sequence flanking the genomic DNA cleavage site (minimal repair matrix) for repairing the cleavage, and includes the correct sequence of the gene for repairing the mutation. Consequently, the targeting construct for gene correction comprises or consists of the

minimal repair matrix; it is preferably from 200 pb to 6000 pb, more preferably from 1000 pb to 2000 pb.

For restoring a functional gene, cleavage of the gene occurs 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 genomic DNA 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.

The subject-matter of the present invention is also the use of at least one meganuclease as defined above, one polynucleotide, preferably included in an expression vector, as defined above for the preparation of a medicament for preventing, improving or curing a disease caused by an infectious agent that presents a DNA intermediate, in an individual in need thereof, said medicament being administrated by any means to said individual.

The subject-matter of the present invention is also a method for preventing, improving or curing a disease caused by an infectious agent that presents a DNA intermediate, 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 subject-matter of the present invention is also the use of at least one meganuclease as defined above, one polynucleotide, preferably included in an expression vector, as defined above, *in vitro*, for inhibiting the propagation, inactivating or deleting an infectious agent that presents a DNA intermediate, in biological derived products or products intended for biological uses or for disinfecting an object.

The subject-matter of the present invention is also a method for decontaminating a product or a material from an infectious agent that presents a DNA intermediate, said method comprising at least the step of contacting a biological

derived product, a product intended for biological use or an object, with a composition as defined above, for a time sufficient to inhibit the propagation, inactivate or delete said infectious agent.

In a particular embodiment, said infectious agent is a virus. For example said virus is an adenovirus (Ad11, Ad21), herpesvirus (HSV, VZV, EBV, CMV, herpesvirus 6, 7 or 8), hepadnavirus (HBV), papovavirus (HPV), poxvirus or retrovirus (HTLV, HIV).

The subject-matter of the present invention is also a composition characterized in that it comprises at least one meganuclease, one polynucleotide, preferably included in an expression vector, as defined above.

In a preferred embodiment of said composition, it comprises a targeting DNA construct comprising the sequence which repairs the site of interest flanked by sequences sharing homologies with the targeted locus 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 encoding the meganuclease, as defined in the present invention.

The subject-matter of the present invention is also products containing at least a meganuclease, one expression vector encoding said meganuclease, and a vector including a targeting construct, as defined above, as a combined preparation for simultaneous, separate or sequential use in the prevention or the treatment of a genetic disease.

For purposes of therapy, the meganuclease 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.

In one embodiment of the uses according to the present invention, the meganuclease is substantially non-immunogenic, i.e., engenders 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 meganuclease can be used either as a polypeptide or as a polynucleotide construct/vector encoding said polypeptide. It is introduced into cells, *in vitro*, *ex vivo* or *in vivo*, 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. 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.

The meganuclease (polypeptide) may be advantageously associated with: liposomes, polyethyleneimine (PEI), and/or 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 the latter case, the sequence of the meganuclease fused with the sequence of a membrane translocating peptide (fusion protein).

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.

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

The uses of the meganuclease and the methods of using said meganucleases according to the present invention include also the use of the polynucleotide, vector, cell, transgenic plant or non-human transgenic mammal encoding said meganuclease, as defined above.

According to another advantageous embodiment of the uses and methods according to the present invention, said meganuclease, polynucleotide, vector, cell, transgenic plant or non-human transgenic mammal are associated with a targeting DNA construct as defined above. Preferably, said vector encoding the meganuclease, comprises the targeting DNA construct, as defined above.

Methods for making I-*Cre*I variants having novel cleavage specificity have been described previously (Epinat *et al.*, Nucleic Acids Res., 2003, 31, 2952-2962; Chames *et al.*, Nucleic Acids Res., 2005, 33, e178, and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458; Smith *et al.*, Nucleic Acids Res., 2006, 34, e149 ; Arnould *et al.*, J. Mol. Biol., Epub 10 May 2007; International PCT Applications WO 2004/067736, WO 2006/097784, WO 2006/097853, WO 2007/049095, WO 2007/057781, WO 2007/060495, WO 2007/049156, WO 2007/093836 and WO 2007/093918). The single-chain meganuclease of the invention may be derived from I-*Cre*I or functional variants thereof by using well-known recombinant DNA and genetic engineering techniques. For example, a sequence comprising the linker coding sequence in frame with, either the 3' end of the N-terminal domain coding sequence or the 5' end of the C-terminal domain coding sequence is amplified from a DNA template, by polymerase chain reaction with specific primers. The PCR fragment is then cloned in a vector comprising the sequence encoding the other domain by using appropriate restriction sites. The single-chain meganuclease as defined in the present invention is produced by expressing the polypeptide as defined above; preferably said

polypeptide is expressed in a host cell or a transgenic animal/plant modified by one expression vector, under conditions suitable for the expression of the polypeptide, and the single-chain meganuclease 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).

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 single-chain I-*Cre*I meganuclease of the present invention and its uses according to the invention, as well as to the appended drawings in which:

- figure 1 is a Cα ribbon representation of I-*Cre*I.

- figure 2 is a schematic representation of the human XPC gene (GenBank accession number NC_000003). The XPC exons are boxed. The XPC4.1 (or C1: SEQ ID NO: 26) sequence (position 20438) is situated in Exon 9.

- figure 3 represents 22 bp DNA targets cleaved by I-*Cre*I or some of its derived variants (SEQ ID NO: 21, 27 to 29, 26, 30 and 31 respectively). C1221 is the I-*Cre*I target. 10GAG_P, 10GTA_P and 5TCT_P are palindromic targets, which differ from C1221 by the boxed motifs. C1 is the XPC target; C3 and C4 are palindromic targets, which are derived respectively from the left and the right part of C1. As shown in the Figure, the boxed motifs from 10GAG_P, 10GTA_P and 5TCT_P are found in the C1 target.

- figure 4 illustrates the C1 target cleavage by heterodimeric combinatorial mutants. The figure displays secondary screening of combinations of C3 and C4 cutters with the C1 target. The H33 mutant is among seven C3 cutters, X2 is among eight C4 cutters and the cleavage of the C1 target by the X2/H33 heterodimer is circled in black.

- figure 5 represents the pCLS0542 meganuclease expression vector map. pCLS0542 is a 2 micron-based replicative vector marked with a LEU2 auxotrophic gene, and an inducible Gal10 promoter for driving the expression of the meganuclease.

- figure 6 represents the map of pCLS1088, a plasmid for expression of meganucleases in mammalian cells.

- figure 7 represents the pCLS1058 reporter vector map. The reporter vector is marked with blasticidine and ampicilline resistance genes. The LacZ tandem repeats share 800 bp of homology, and are separated by 1.3 kb of DNA. They are surrounded by EF1-alpha promoter and terminator sequences. Target sites are cloned using the Gateway protocol (INVITROGEN), resulting in the replacement of the CmR and ccdB genes with the chosen target site

- figure 8 illustrates the yeast screening of the eighteen single chain constructs against the three XPC targets C1, C3 and C4. Each single chain molecule is referred by its number described in Table I. For each four dots yeast cluster, the two left dots are the result of the experiment, while the two right dots are various internal controls to assess the experiment quality and validity.

- figure 9 illustrates the cleavage of the C1, C3 and C4 XPC targets by the two X2-L1-H33 and X2-RM2-H33 single chain constructs in an extrachromosomal assay in CHO cells. Background corresponds to the transfection of the cells with an empty expression vector. Cleavage of the S1234 I-*Sce*I target by I-*Sce*I in the same experiment is shown as a positive control.

- figure 10 illustrates the cleavage of the C1, C3 and C4 XPC targets by the X2/H33 heterodimer and the two X2-L1-$H33_{G19S}$ and X2-RM2-$H33_{G19S}$ single chain constructs in an extrachromosomal assay in CHO cells. Background corresponds to the transfection of the cells with an empty expression vector. Cleavage of the S1234 I-*Sce*I target by I-*Sce*I in the same experiment is shown as a positive control.

- figure 11 illustrates the yeast screening of three XPC single chain molecules X2-L1-H33, SCX1 and SCX2 against the three XPC targets (C1, C3 and C4). SCX1 is the X2(K7E)-L1-H33(E8K,G19S) molecule and SCX2 stands for the X2(E8K)-L1-H33(K7E,G19S) molecule. For each four dots yeast cluster, the two left dots are the result of the experiment, while the two right dots are various internal controls to assess the experiment quality and validity.

- figure 12 is a schematic representation of the *Cricetulus griseus* Hypoxanthine-Guanine Phosphoribosyl Transferase (HPRT) mRNA (GenBank accession number J00060.1). The ORF is indicated as a grey box. The HprCH3 target site is indicated with its sequence (SEQ ID NO: 34) and position.

- figure 13 represents 22 bp DNA targets cleaved by I-*Cre*I or some of its derived variants (SEQ ID NO: 21, 27 and 32 to 36, respectively). C1221 is the I-*Cre*I target. 10GAG_P, 10CAT_P and 5CTT_P are palindromic targets, which differ from C1221 by the boxed motifs. HprCH3 is the HPRT target, HprCH3.3 and HprCH3.4 are palindromic targets, which are derived respectively from the left and the right part of HprCH3. As shown in the Figure, the boxed motifs from 10GAG_P, 10CAT_P and 5CTT_P are found in the HprCH3 target.

- figure 14 illustrates the yeast screening of the MA17 and H33 homodimer, of the HPRT heterodimer and of three HPRT single chain molecules against the three HPRT targets HprCH3, HprCH3.3 and HprCH3.4. H is the MA17 / H33 heterodimer. Since it results from co-expression of MA17 and H33, there are

actually three molecular species in the yeast: the two MA17 and H33 homodimers, together with the MA17/H33 heterodimer. Homodimer formation accounts for cleavage of the HprCH3.3 and HprCH3.4 targets. SC1 to SC3 are MA17-L1-H33, MA17-L1-$H33_{G19S}$ and MA17-RM2-H33, respectively. For each four dots yeast cluster, the two left dots are the result of the experiment, while the two right dots are various internal controls to assess the experiment quality and validity.

- figure 15 illustrates the cleavage of the HprCH3, HprCH3.3 and HprCH3.4 HPRT targets by the MA17/H33 heterodimer and the four HPRT single chain constructs (MA17-L1-H33, MA17-L1-$H33_{G19S}$, MA17-RM2-H33 and MA17-RM2-$H33_{G19S}$) in an extrachromosomal assay in CHO cells. Background corresponds to the transfection of the cells with an empty expression vector. Cleavage of the S1234 I-*Sce*I target by I-*Sce*I in the same experiment is shown as a positive control.

- figure 16 is a schematic representation of the human RAG1 gene (GenBank accession number NC_000011). Exonic sequences are boxed, and the Exon-Intron junctions are indicated. ORF is indicated as a grey box. The RAG1.10 sequence is indicated with its sequence (SEQ ID NO: 41) and position.

- figure 17 represents 22 bp DNA targets cleaved by I-*Cre*I or some of its derived variants (SEQ ID NO: 21 and 37 to 43, respectively). C1221 is the I-*Cre*I target. 10GTT_P, 5CAG_P, 10TGG_P and 5GAG_P are palindromic targets, which differ from C1221 by the boxed motifs. RAG1.10 is the RAG1 target, RAG1.10.2 and RAG1.10.3 are palindromic targets, which are derived from the left and the right part of RAG1.10, respectively. As shown in the Figure, the boxed motifs from 10GTT_P, 5CAG_P, 10TGG_P and 5GAG_P are found in the RAG1.10 target.

- figure 18 illustrates the yeast screening of the four RAG1 single chain molecules against the three RAG1 targets RAG1.10, RAG1.10.2 and RAG1.10.3. SC1 to SC4 represent M2-L1-M3, $M2_{G19S}$-L1-H33, M2-RM2-H33 and $M2_{G19S}$-RM2-M3, respectively. Activity of the M2 and M3 I-*Cre*I mutants and the M2/M3 heterodimer against the three RAG1 targets is also shown. For each four dots yeast cluster, the two left dots are the result of the experiment, while the two right dots are various internal controls to assess the experiment quality and validity.

- figure 19 illustrates the yeast screening of two single chain molecules SC1 and SC2 against the three RAG1.10 targets. SC1 is the M3-RM2-M2

molecule and SC2 stands for the M3(K7E K96E)-RM2-M2(E8K E61R) molecule. For each 4 dots yeast cluster, the two left dots are the result of the experiment, while the two right dots are various internal controls to assess the experiment quality and validity.

Example 1: The making of a single chain I-*Cre*I derived meganuclease cleaving the human XPC gene.

Xeroderma Pigmentosum (XP) is a rare autosomal recessive genetic disease characterized by a hypersensitivity to exposure to ultraviolet (UV) rays and a high predisposition for developing skin cancers. The human XPC gene involved in Xeroderma Pigmentosum was scanned for potential target sequences. A potential 22bp DNA target that was called C1 (cgagatgtcacacagaggtacg; SEQ ID NO: 26), was localized at the end of the XPC ninth exon (Figure 2). The engineering of I-*Cre*I derived mutants able to cleave the C1 target has been described previously (Arnould *et al.*, J. Mol. Biol., Epub 10 May 2007; International PCT Application WO 2007/093836 and WO 2007/093918). Briefly, the C1 sequence was divided into two palindromic half-targets called C3 and C4 (Figure 3). As the C3 target is identical to the 10GAG_P target but with a single difference at position ±6, I-*Cre*I derived mutants able to cleave the 10GAG_P target were screened against the C3 target. The mutant H33 bearing the single mutation Y33H (substitution at position 33 of a tyrosine by a histidine residue) in comparison to the I-*Cre*I wild-type sequence, was isolated as a strong C3 cutter. The C4 target is a combination of the 10GTA_P and 5TCT_P targets. I-*Cre*I mutants able to cleave the 10GTA_P and I-*Cre*I mutants able to cleave the 5TCT_P target were combined and screened against the C4 DNA target, as described previously (Smith *et al.*, Nucleic Acids Research, 2006, 34, e149; International PCT Applications WO 2007/049095 and WO 2007/057781). The I-*Cre*I mutant called X2 was isolated as a strong C4 cutter. The X2 mutant bears the following mutations in comparison with the I-*Cre*I wild type sequence: Y33H, Q38A, S40Q, Q44K, R68Q, R70S and D75N. The last step consisted in the yeast co-expression of the H33 and X2 I-*Cre*I mutants, as described previously (International PCT Application WO 2006/097854 and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458), which resulted in the strong cleavage of the XPC C1 DNA target (Figure 4).

The X2/H33 XPC heterodimer obtained by coexpression of the two mutants cleaves the C1 target but also the C3 and C4 targets, because of the presence of the two homodimers. To avoid these side effects, a new way for designing a single chain molecule composed of the two I-*Cre*I derived mutants X2 and H33 was conceived. For that purpose, a full length X2 N-terminal mutant was maintained in the single chain design, andseveral constructs of the type X2-L-H33, where L is a protein linker, were engineered. In this nomenclature, X2 will be referred as the N-terminal mutant and H33 as the C-terminal mutant. Another important issue was the sequence identity of the two mutants in the single chain molecule. In fact, the two I-*Cre*I mutants X2 and H33 have almost the same nucleic sequence, which raises the problem of the stability of such a construct with the risk of a recombination event between the two almost identical sequences. To avoid or at least reduce this possibility, another I-*Cre*I version, called I-CreI CLS, was used to code for the H33 mutant. The nucleic sequence of I-*Cre*I CLS (SEQ ID NO: 24) has been rewritten from the I-*Cre*I wild type sequence (I-CreI wt; SEQ ID NO: 22) using the codon usage and the genetic code degeneracy. It means that I-CreI CLS shares 73% nucleic sequence identity with I-*Cre*I wt and has three single amino acid mutations (T42A, E110W and Q111R), which do not alter I-*Cre*I activity. The Y33H was then introduced in the I-*Cre*I CLS version. Activity of the H33 CLS mutant was checked against the C3 target and was shown to be as strong as for the H33 mutant in the I-CreI wt version.

Using the H33 CLS mutant, 18 single chain versions of the type X2-L-H33 were built, where L is a linker, different for each of the 18 versions. The G19S mutation was also introduced in the C-terminal H33 mutant for two single chain molecules. Activity of these different single chain constructs against the C1 XPC target and its two derivatives C3 and C4 was then monitored in yeast and, for some of them, in CHO cells using an extrachromosomal assay.

1) Material and Methods

a) Introduction of the Y33H mutation into the I-*Cre*I CLS version

Two overlapping PCR reactions were performed using two sets of primers: Gal10F (5'-gcaactttagtgctgacacatacagg-3'; SEQ ID NO: 44) and H33Revp60 (5'-ctggtgtttgaacttgtgagattgatttggttt-3'; SEQ ID NO: 45) for the first fragment and H33Forp60 (5'-aaaccaaatcaatctcacaagttcaaacaccag-3'; SEQ ID NO: 46) and Gal10R

(5'-acaaccttgattggagacttgacc-3'; SEQ ID NO: 47) for the second fragment. Approximately 25 ng of each PCR fragment and 75 ng of vector DNA pCLS0542 (Figure 5) linearized by digestion with *Nco*I and *Eag*I were used to transform the yeast *Saccharomyces cerevisiae* strain FYC2-6A (MATα, trp1Δ63, leu2Δ1, his3Δ200) using a high efficiency LiAc transformation protocol (Gietz, R. D. and Woods R. A., Methods Enzymol., 2002, 350, 87-96). An intact coding sequence containing the Y33H mutation is generated by *in vivo* homologous recombination in yeast.

b) Sequencing of mutants

To recover the mutant expressing plasmids, yeast DNA was extracted using standard protocols and used to transform *E. coli*. Sequencing of mutant ORF was then performed on the plasmids by MILLEGEN SA.

c) Cloning of the eighteen XPC single chain molecules

Eighteen independent PCR reactions were performed on the H33 mutant in the I-*Cre*I CLS version. Each PCR reaction uses the same reverse primer CreCterR60 (5'-tagacgagctcctaaggagaggacttttttcttctcag-3'; SEQ ID NO: 48) and a specific forward primer. The eighteen forward primers that were used are :

- L1EagI: 5'-tatcggccggtggcggaggatctggcggcggtggatccggtggtggaggctccggaggaggtggctctaacaaagagttcctgctgtatcttgctgga-3' (SEQ ID NO: 49)
- YPP: 5'-tatcggccggtaaatcttccgattccaagggtattgatctgactaatgttactctgcctgataccctacttattccaaagctgcctctgatgctattcctccagctaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 50)
- AOL: 5'-tatcggccggtctggagtatcaggctccttactcttcccctccaggtcctccttgttgctccggttcctctggctcctctgctggttgttctaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 51)
- CXT:5'-tatcggccggtctgtcctatcattattctaatggtggctcccctacttctgatggtccagctctgggtggcatttctgatggtggcgctactaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 52),
- BQY: 5'-tatcggccggtgattcctctgtttctaattccgagcacattgctcctctgtctctgccttcctctcctccatctgttggttctaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 53)
- VSG: 5'-tatcggccggtgcttctcagggttgtaaacctctggctctgcctgagctgcttactgaggattcttataatactgataacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 54)
- BYM: 5'-tatcggccggtaatcctattcctggtctggatgagctgggtgttggcaactctgatgctgccgctcctggcactaacaaagagttcctgctgtatcttgctggattt-3' (SEQ ID NO: 55)
- MCJ: 5'-tatcggccggtgctcctactgagtgttctccttccgctctgacccagcctccatccgcttctggttccctgaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 56)
- GSmid: 5'-tatcggccggtggaggcggttctggaggcggtggctctggtggaggcggttccggtggaggcggatctggtggaggcggttctaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 57)
- GSshort: 5'-tatcggccggtggaggcggttctggaggcggtggctctggtggaggcggttccaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 58)
- GSxshort: 5'-tatcggccggtggaggcggttctggaggcggtggctctaacaaagagttcctgctgtatcttgctggattt-3(SEQ ID NO: 59)

- PPR: 5'-tatcggccggtcaggttacttctgctgccggtcctgctactgttccatctggtaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 60)

-SBA1: 5'-tatcggccggtggatctcctctgaagccttctgccccaaagattcctataggtggctccaacaaagagttcctgctgtatcttgctgg attt-3'(SEQ ID NO: 61)

- SBA2: 5'-tatcggccggtggatctcctctgaagccttctgccccaaagattcctataggtggctccccactgaaaccttccgcacctaaaatcccaattggt ggctctaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 62)

- LP1: 5'-tatcggccggtggatctcctctgtctaaaccaattccaggcggttccaacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 63)

- LP2: 5'-tatcggccggtggatctcctctgtctaaaccaattccaggcggttccccactgtcaaagccaatccctggcggttctaacaaagagttcctgctgtatct tgctggattt-3'(SEQ ID NO: 64)

- RM1: 5'-tatcggccggtggatctgataagtataatcaggctctgtctgagcgtcgcgcctacgttgtcgccaataacctggtttccggtggaggcggttccaac aaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 65)

- RM2: 5'-tatcggccggtggatctgataagtataatcaggctctgtctaaatacaaccaagcactgtccaagtacaatcaggccctgtctggtggaggcggttc caacaaagagttcctgctgtatcttgctggattt-3'(SEQ ID NO: 66).

All PCR fragments were purified and digested by *Eag*I and *Sac*I and each PCR fragment was ligated into the yeast expression vector for the X2 mutant also digested with *Eag*I and *Sac*I. After sequencing of the clones, all single chain molecules in the yeast expression vector were obtained.

d) Introduction of the G19S mutation into the C-terminal H33 mutant of the two single chain molecules X2-L1-H33 and X2-RM2-H33

Two overlapping PCR reactions were performed using two sets of primers: Gal10F (5'-gcaactttagtgctgacacatacagg-3': SEQ ID NO: 44) and G19SRev (5'gcaatgatggagccatcagaatccacaaatccagc-3': SEQ ID NO: 67) for the first fragment and G19SFor60 (5'-gctggatttgtggattctgatggctccatcattgc-3': SEQ ID NO: 68) and Gal10R (5'-acaaccttgattggagacttgacc-3': SEQ ID NO: 47) for the second fragment. Approximately 25 ng of each PCR fragment and 75 ng of vector DNA (pCLS0542; Figure 5) linearized by digestion with *Nco*I and *Eag*I were used to transform the yeast *Saccharomyces cerevisiae* strain FYC2-6A (MATα, trp1Δ63, leu2Δ1, his3Δ200) using a high efficiency LiAc transformation protocol (Gietz R. D. and Woods R. A., Methods Enzymol., 2002, 350, 87-96). An intact coding sequence containing the G19S mutation is generated by *in vivo* homologous recombination in yeast.

e) Introduction of the K7E, E8K and G19S mutations in the XPC X2-L1-H33 Single Chain molecule

First, the G19S mutation was introduced in the X2-L1-H33 molecule. Two overlapping PCR reactions were performed on the single chain molecule cloned in the pCLS0542 yeast expression vector. The first PCR reaction uses the primers: Gal10F (5'-gcaactttagtgctgacacatacagg-3'; SEQ ID NO: 44) and G19SRev60 (5'-

gcaatgatggagccatcagaatccacaaatccagc-3'; SEQ ID NO: 67) and the second PCR reaction, the primers G19SFor60 (5'-gctggatttgtggattctgatggctccatcattgc-3; SEQ ID NO: 68) and Gal10R (5'-acaaccttgattggagacttgacc-3'; SEQ ID NO: 47). Approximately 25 ng of each PCR fragment and 75 ng of vector DNA (pCLS0542; Figure 5) linearized by digestion with *Nco*I and *Eag*I were used to transform the yeast Saccharomyces cerevisiae strain FYC2-6A (MATα, trp1Δ63, leu2Δ1, his3Δ200) using a high efficiency LiAc transformation protocol (Gietz R. D. and Woods R. A., Methods Enzymol., 2002, 350, 87-96). An intact coding sequence containing the G19S mutation was generated by *in vivo* homologous recombination in yeast.

In a second step, the K7E and E8K mutations were introduced in the X2-L2-H33(G19S) molecule by performing three overlapping mutations. For the SCX1 molecule, the 3 PCR reactions use three primers set, which are respectively: Gal10F and K7ERev (5'-gtacagcaggaactcttcgttatatttggtattgg-3'), K7EFor (5'-aataccaaatataacgaagagttcctgctgtacc-3'; SEQ ID NO: 69) and E8KRevSC (5'-aagatacagcaggaacttttttgttagagccacc-3'; SEQ ID NO: 70), E8KForSc (5'-ggtggctctaacaaaaagttcctgctgtatctt-3'; SEQ ID NO: 71) and Gal10R. For the SCX2 molecule, the 3 PCR reactions use three primers set, which are respectively: Gal10F and E8KRev (5'-caggtacagcaggaacttttttgttatatttgg-3'; SEQ ID NO: 72), E8KFor (5'-accaaatataacaaaaagttcctgctgtacctgg-3'; SEQ ID NO: 73) and K7ERevSC (5'-aagatacagcaggaactcttcgttagagccacc-3'; SEQ ID NO: 74), K7EForSc (5'-ggtggctctaacgaagagttcctgctgtatctt-3'; SEQ ID NO: 75) and Gal10R. For both constructs, approximately 25ng of each PCR fragment and 75ng of vector DNA (pCLS0542; Figure 5) linearized by digestion with NcoI and EagI were used to transform the yeast Saccharomyces cerevisiae strain FYC2-6A (MATα, trp1Δ63, leu2Δ1, his3Δ200) using a high efficiency LiAc transformation protocol (Gietz R D and Woods R A Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. 2002; 350:87-96). An intact coding sequence for the SCX1 or SCX2 constructs was generated by in vivo homologous recombination in yeast.

f) Mating of meganuclease expressing clones and screening in yeast

Screening was performed as described previously (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, and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458). Mating was performed using a colony gridder (QpixII, Genetix). Mutants were gridded on nylon filters covering YPD plates, using a low gridding density (about 4 spots/cm2). A second gridding process was performed on the same filters to spot a second layer consisting of different reporter-harboring yeast strains for each target. Membranes were placed on solid agar YPD rich medium, and incubated at 30 °C for one night, to allow mating. Next, filters were transferred to synthetic medium, lacking leucine and tryptophan, with galactose (2 %) as a carbon source, and incubated for five days at 37 °C, to select for diploids carrying the expression and target vectors. After 5 days, filters were placed on solid agarose medium with 0.02 % X-Gal in 0.5 M sodium phosphate buffer, pH 7.0, 0.1 % SDS, 6 % dimethyl formamide (DMF), 7 mM β-mercaptoethanol, 1 % agarose, and incubated at 37 °C, to monitor β-galactosidase activity. Results were analyzed by scanning and quantification was performed using appropriate software.

g) Cloning of the XPC Single Chain constructs into a mammalian expression vector

Each mutant ORF was amplified by PCR using the primers CCM2For (5'-aagcagagctctctggctaactagagaacccactgcttactggcttatcgaccatggccaatacc aaatataacaaagagttcc-3': SEQ ID NO: 76) and CCMRev60 (5'-ctgctctagactaaggagaggacttttttcttctcag-3': SEQ ID NO: 77). The PCR fragment was digested by the restriction enzymes *Sac*I and *Xba*I, and was then ligated into the vector pCLS1088 (Figure 6) digested also by *Sac*I and *Xba*I. Resulting clones were verified by sequencing (MILLEGEN).

h) Cloning of the C1, C3 and C4 targets in a vector for extrachromosomal assay in CHO cells

The target of interest was cloned as follow: oligonucleotide corresponding to the target sequence flanked by gateway cloning sequence was ordered from PROLIGO. 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, Figure 7).

i) Extrachromosomal assay in CHO cells

CHO cells were transfected with Polyfect transfection reagent according to the supplier's (QIAGEN) protocol. 72 hours after transfection, culture

medium was removed and 150 µl of lysis/revelation buffer added for β-galactosidase liquid assay (1 liter of buffer contained 100 ml of lysis buffer (Tris-HCl 10 mM pH7.5, NaCl 150 mM, Triton X100 0.1 %, BSA 0.1 mg/ml, protease inhibitors), 10 ml of Mg 100X buffer ($MgCl_2$ 100 mM, β-mercaptoethanol 35 %), 110 ml ONPG 8 mg/ml and 780 ml of sodium phosphate 0.1M pH7.5). After incubation at 37 °C, OD was measured at 420 nm. The entire process was performed on an automated Velocity11 BioCel platform.

2) Results

a) Cleavage activity of the 18 XPC single-chain meganucleases against the three XPC targets

Table I shows the different linkers that have been used to build the different single chain molecules. For each single chain molecule, the linker begins after the last residue (P163) of the N-terminal X2 mutant and after the linker, the N-terminal H33 mutant begins at residue N6.

Table I: Linkers

Number	Linker Name	Size (aa)	SEQ ID NO:	Primary Sequence
1	L1	22	3	-AA$(GGGGS)_4$-
2	YPP	35	4	-AAGKSSDSKGIDLTNVTLPDTPTYSKAASDAIPPA-
3	AOL	30	5	-AAGLEYPQAPYSSPPGPPCCSGSSGSSAGCS-
4	CXT	30	6	-AAGLSYHYSNGGSPTSDGPALGGISDGGAT-
5	BQY	27	7	-AAGDSSVSNSEHIAPLSLPSSPPSVGS-
6	VSG	25	8	-AAGASQGCKPLALPELLTEDSYNTD-
7	BYM	24	9	-AAGNPIPGLDELGVGNSDAAAPGT-
8	MCJ	23	10	-AAGAPTECSPSALTQPPSASGSL-
9	GSmid	27	11	-AA$(GGGGS)_5$-
10	Gsshort	17	12	-AA$(GGGGS)_3$-
11	GSxshort	12	13	-AA$(GGGGS)_2$-
12	PPR	17	14	-AAGQVTSAAGPATVPSG-
13	SBA1	19	15	-AAGGSPLKPSAPKIPIGGS-
14	SBA2	33	16	-AAGGSPLKPSAPKIPIGGSPLKPSAPKIPIGGS-
15	LP1	15	17	-AAGGSPLSKPIPGGS-
16	LP2	25	18	-AAGGSPLSKPIPGGSPLSKPIPGGS-
17	RM1	31	19	-AAGGSDKYNQALSERRAYVVANNLVSGGGGS-
18	RM2	32	2	-AAGGSDKYNQALSKYNQALSKYNQALSGGGGS-

The cleavage activity of the 18 XPC single chain molecules was monitored against the three XPC targets C1, C3 and C4, using the yeast screening assay previously described (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, and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458). As shown in Figure 8, all the single chain constructs cleave the C1 target, but also strongly the C3 target while activity toward the C4 target is also detectable. The two single chain molecules X2-L1-H33 and X2-RM2-H33 were cloned into a mammalian expression vector and their activity toward the three XPC targets was checked in CHO cells using an extrachromosomal assay (Figure 9). This assay confirmed the yeast cleavage profile (Figure 9). Furthermore, X2-RM2-H33 is more active toward the C1 target than X2-L1-H33 in CHO cells. C4 cleavage is barely detectable in CHO cells. No cleavage of C3 was observed with X2, and similarly, H33 does not cleave C4 (data not shown). Thus, the strong cleavage of the C3 target with single chain molecules suggests that the linker does not abolish the formation of intermolecular species, resulting from interaction between the dimerization interfaces of the H33 units from two distinct molecules. The formation of pseudo H33 homodimers would then be responsible for C3 cleavage.

b) Effect of the G19S mutation alone on the specificity of cleavage of the XPC single-chain meganuclease

In order to test this hypothesis, the G19S mutation (was introduced in the H33 C-terminal mutant of the two single chain molecules. The G19S mutation (mutation of residue 19 from I-*Cre*I, according to pdb 1G9Y numeration) has been shown before to abolish the formation of functional homodimers while enhancing the activity of the hetrodimers displaying a G19S monomer (International PCT Application WO 2008/010093). The two single chain molecules that were obtained (X2-L1-H33$_{G19S}$ and X2-RM2-H33$_{G19S}$), were then profiled against the three XPC targets using the extrachromosomal assay in CHO cells. Figure 10 shows that the G19S mutation does not only increases the activity toward the C1 target but also greatly reduces the activity toward the C3 target. The profile cleavage of the X2/H33 heterodimer against the three XPC targets is also shown on Figure 10, for comparison.

These results confirm the hypothesis of intermolecular species formation, resulting from the interaction of two H33 units. Similarly, interaction between two X2 units probably account for weak cleavage of the C4 target (Figures 8 and 9).Thus, although the introduction of a linker between the X2 and H33 monomers

might favour intramolecular interactions (resulting in reduced cleavage of C4 for example), it does not abolish intermolecular interactions.

Nevertheless, the engineered XPC single chain molecule X2-RM2-$H33_{G19S}$ cleaves more strongly the C1 XPC target than the X2/H33 heterodimer and has much reduced cleavage activities toward the two palindromic C3 and C4 targets than the same heterodimer. The single chain molecule X2-RM2-$H33_{G19S}$ displays a much better specificity than the X2/H33 heterodimer and has an activity level comparable to that of I-*Sce*I, the gold standard in the field of homologous recombination induced by DNA double strand break.

c) Effect of the combination of the G19S mutation with another mutation that impairs the formation of a functional homodimer on the specificity of cleavage of the XPC single- chain meganuclease

Figure 11 shows the activity of the three single chain molecules X2-L1-H33, SCX1 and SCX2 against the three XPC targets in a yeast screening assay. The initial single chain molecule has a strong cleavage activity against the C1 and C3 target but introduction of the K7E/E8K and G19S mutations to generate the SCX1 and SCX2 molecules promotes an increased cleavage activity toward the C1 target and a complete abolition of the cleavage activity toward the C3 target. Thus, the mutations K7E/E8K and G19S can be successfully introduced in a single chain molecule to improve its specificity without affecting its cleavage activity toward the DNA target of interest.

Example 2: The making of a single chain I-CreI derived meganuclease cleaving the *Cricetulus griseus* HPRT gene.

The Hypoxanthine-Guanine Phosphorybosyl Transferase (HPRT) gene from *Cricetulus griseus* was scanned for a potential target site. A 22bp sequence called HprCH3 (cgagatgtcatgaaagagatgg: SEQ ID NO: 34) was identified in the gene ORF (Figure 12). Two derived palindromic targets HprCH3.3 and HprCH3.4 were derived from the HprCH3 target (Figure 13). As the HprCH3.3 target is identical to the C3 target described above in Example 1, the H33 I-*Cre*I mutant is able to cleave strongly HprCH3.3 (C3). The HprCH3.4 target is a combination of the 10CAT_P and 5CTT_P targets. I-*Cre*I mutants able to cleave the 10CAT_P target were obtained as previously described in Smith *et al.*, Nucleic Acids Research, 2006, 34, e149;

International PCT Applications WO 2007/049156 and WO 2007/060495 and I-*Cre*I mutants able to cleave the 5CTT_P target were obtained as previously described in Arnould *et al.*, J Mol. Biol., 2006; 355, 443-458; International PCT Applications WO 2006/097784 and WO 2006/097853. The mutants were combined as previously described in Smith *et al.*, Nucleic Acids Research, 2006, 34, e149; International PCT Applications WO 2007/049095 and WO 2007/057781 and then screened against the HprCH3.4 DNA target. The I-*Cre*I mutant called MA17 was isolated as a HprCH3.4 cutter. However, it was found to cleave also the HprCH3.3 target, due to a relaxed specificity (see Figure 14). The MA17 mutant bears the following mutations in comparison with the I-*Cre*I wild type sequence: S32T, Y33H, Q44R, R68Y, R70S, S72T, D75N and I77N. The last step consisted in the yeast co-expression of the H33 and MA17 I-*Cre*I mutants, as described previously (International PCT Application WO 2006/097854 and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458), which resulted in the cleavage of the HprCH3 DNA target.

Two HPRT single chain constructs were engineered following the same scheme as in example 1. The two L1 and RM2 linkers (see Table I of Example 1) were used, resulting in the production of the MA17-L1-H33 and MA17-RM2-H33 single chain molecules. In a second step, the G19S mutation was introduced in the C-terminal H33 mutant, resulting in a two other single chain meganuclease, MA17-L1-$H33_{G19S}$ and MA17-RM2-$H33_{G19S}$. The activity of these different constructs was then monitored in yeast and in CHO cells against the HprCH3 target and its two derivatives HprCH3.3 and HprCH3.4 targets.

1) Material and Methods

See example 1

2) Results

The activity of three single chain molecules (MA17-L1-H33, MA17-L1-$H33_{G19S}$ and MA17-RM2-H33) against the three HPRT targets HprCH3, HprCH3.3 and HprCH3.4 was monitored using the previously described yeast assay (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, and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458). As shown in Figure 14, MA17-L1-H33 does not cleave the HprCH3 target, which is cleaved by the two other single chain

molecules. Thus, the RM2 linker seems to be better adapted to the way we have engineered the single chain constructs, as already observed in example 1. These results also confirm that the presence of the G19S mutation enhances the heterodimeric activity. All three single chain molecules cleave very strongly the HprCH3.3 (identical to the C3 target from example 1), but do not cleave the HprCH3.4, in contrast with the MA17/H33 heterodimer. In this case, cleavage of HprCH3.3 (C3) is not necessarily due to the formation of intermolecular species: since the MA17 and H33 mutants both cleave the HprCH3.3 target as homodimers, a MA17/H33 heterodimer or a MA17-RM2-H33 single chain monomer could in principle cleave HprCH3.3. This hypothesis is confirmed by the persistent cleavage of HprCH3.3 by MA17-L1-H33$_{G19S}$ and MA17-RM2-H33$_{G19S}$. Next, four single chain molecules (MA17-L1-H33, MA17-L1-H33$_{G19S}$, MA17-RM2-H33 and MA17-RM2-H33$_{G19S}$) were cloned into a mammalian expression vector tested in CHO cells using for cleavage of the three HPRT targets (Figure 15). The MA17-RM2- H33$_{G19S}$ single chain molecule displayed the strongest activity. Again strong cleavage of HprCH3.3 (C3) was observed, while cleavage of HprCH3.4 was low or absent.

Example 3: The making of a single chain I-*Cre*I derived meganuclease cleaving the human RAG1 gene.

RAG1 is a gene involved in the V(D)J recombination process, which is an essential step in the maturation of immunoglobulins and T lymphocytes receptors (TCRs). Mutations in the RAG1 gene result in defect in lymphocytes T maturation, always associated with a functional defect in lymphocytes B, which leads to a Severe Immune Combined Deficiency (SCID). A 22 bp DNA sequence located at the junction between the intron and the second exon of the human RAG1, called RAG1.10 (SEQ ID NO: 41; Figure 16) had been identified as a potential cleavable sequence by our meganucleases (Smith *et al.*, Nucleic Acids Research, 2006, 34, e149 International PCT Application WO 2008/010093). The RAG1.10 sequence was derived into two palindromic RAG1.10.2 and RAG1.10.3 sequences (Figure 17). RAG1.10.2 target is a combination of the 10GTT_P and 5CAG_P targets and RAG1.10.3 target is a combination of the 10TGG_P and 5GAG_P targets. Strong cutters for both RAG1.10.2 and RAG1.10.3 targets were obtained by combining I-*Cre*I mutants able to cleave the 10GTT_P and 5CAG_P targets from one side, and the

10TGG_P and 5GAG_P targets from the other side, as described previously in Smith *et al.*, Nucleic Acids Research, 2006, 34, e149; International PCT Applications WO 2007/049095, WO 2007/057781, WO 2007/049156, WO 2007/060495, WO 2006/097784 and WO 2006/097853; Arnould *et al.*, J Mol. Biol., 2006; 355, 443-458). Coexpression of the cutters as described previously (International PCT Application WO 2006/097854 and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458) leads then to a strong cleavage of the RAG1.10 target. The M2/M3 RAG1.10 heterodimer gives the strongest cleavage in yeast (International PCT Application WO 2008/010093). M2 is a RAG1.10.2 cutter and bears the following mutations in comparison with the I-*Cre*I wild type sequence: N30R, Y33S, Q44A, R68Y, R70S and I77R. M3 is a RAG1.10.3 cutter and bears the following mutations in comparison with the I-*Cre*I wild type sequence: K28N, Y33S, Q38R, S40R, Q44Y, R70S, D75Q and I77V.

Following the same experimental scheme as in Examples 1 and 2, three single chain constructs were engineered using the two linkers L1 and RM2 (see Table I of Example 1), resulting in the production of the three single chain molecules: M2-L1-M3, M2-RM2-M3 and M3-RM2-M2. In a second step, the G19S mutation was introduced in the N-terminal M2 mutant from the M2-L1-M3 and M2-RM2-M3 single chain molecules, resulting in two additional constructs. In addition, mutations K7E, K96E were introduced into the M3 mutant and mutations E8K, E61R into the M2 mutant of M3-RM2-M2 to create the single chain molecule: M3(K7E K96E)-RM2-M2(E8K E61R) that is called further SC_OH The six single chain constructs were then tested in yeast for cleavage of the RAG1.10 target and of its two RAG1.10.2 and RAG1.10.3 derivatives.

1) Material and Methods

See example 1

Cloning of the SC_OH single chain molecule

A PCR reaction was performed on the M2 mutant carrying the K7E and K96E mutations cloned in the pCLS0542 yeast expression vector. The PCR reaction uses the reverse primer CreCterSacI (5'-tagacgagctcctacggggaggatttcttcttctcgct-3'; SEQ ID NO: 78) and the forward primer.

RM2 (5'-tatcggccggtggatctgataagtataatcaggctctgtctaaatacaaccaagcactgtccaagtacaatc aggccctgtctggtggaggcggttccaacaaagagttcctgctgtatcttgctggattt-3'; SEQ ID NO: 66).

The PCR fragment was purified and digested by *Eag*I and *Sac*I and ligated into the yeast expression vector for the M3 mutant carrying the mutations E8K and E61R also digested with *Eag*I and *Sac*I. After sequencing of the clones, a SC_OH single chain molecule was obtained

2) Results

The activity of the four RAG1 single chain molecules (M2-L1-M3, $M2_{G19S}$-L1-M3, M2-RM2-M3 and $M2_{G19S}$-RM2-M3) was monitored against the three RAG1 targets RAG1.10, RAG1.10.2 and RAG1.10.3 (Figure 18) using the previously described yeast assay (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, and Arnould *et al.*, J. Mol. Biol., 2006, 355, 443-458). As observed previously in examples 1 and 2, the RM2 linker seems to be better adapted to the way the single chain constructs were engineered: cleavage of the RAG1.10 target was observed with the M2-RM2-M3 molecule, but not with M2-L1-M3. In addition, M2-RM2-M3 was found to cleave also RAG1.10.2. Since M3 does not cleave RAG1.10.2 (Figure 18), these results suggest that intermolecular interactions can still result in contact between two M2 units, that would form a homodimeric or pseudo-homodimeric species responsible for cleavage of the palindromic RAG1.10.2 target.

Introduction of the G19S mutation in the M2 mutant improved the activity of both molecule, since $M2_{G19S}$-L1-M3 cleaves the RAG1.10 target, and $M2_{G19S}$-RM2-M3 is more active than M2-L1-M3. In addition, the G19S mutation, which has been shown previously to impair formation of functional homodimers (see example 1 of the present Application and International PCT Application WO 2008/010093), abolishes RAG1.10.2 cleavage. This result is consistent with the hypothesis that interaction between M2 units from distinct M2-RM2-M3 single chain molecules is still possible. However, the single chain structure might favour intramolecular interactions in some extent, for in contrast with M3, the M2-RM2-M3 molecule does not cleave RAG1.10.3 In conclusion, the $M2_{G19S}$-RM2-M3 RAG1 single chain molecule cleaves the RAG1.10 target at a saturating level in yeast,

comparable to that observed with the M2/M3 heterodimer, and does not show any cleavage of the two derived palindromic targets RAG1.10.2 and RAG1.10.3.

The yeast screen of the two single chain molecules M3-RM2-M2 and SC_OH against the three RAG1.10 targets depicted in figure 19 shows that introduction of the K7E/E8K and E61R/K96E allows for the abolition of the homodimeric activity against the RAG1.10.2 target without reducing the single chain cleavage activity for the RAG1.10 target. It is therefore possible to introduce these mutations in a single chain molecule to improve its specificity without affecting its activity toward the DNA target of interest.

Example 4: Making of a RAG1 single chain molecule with different N- and C-terminal endings for both subunits.

Using the M2-RM2-M3 RAG1 single chain molecule, new single chain constructs with different N- and C-terminal endings for both subunits were engineered. These new constructs could allow to pinpoint the best possible position of the linker joining the two I-*Cre*I derived mutants. The N-terminal of I-*Cre*I consists in a 6 residues loop followed by the LAGLIDADG α-helix starting at residue K7. In the I-*Cre*I structure (PDB code 1G9Y), the last 10 C-terminal residues are not visible because probably disordered. In the structure, the C-terminus ends at residue D153 with the helix α6 covering residues 145 to 150. So several single chain constructs were made where the N-terminus of the M2 or M3 mutant begins with the residue M1, N2 or N6 and C-terminus ends at different positions, respectively S145, L152, S156 and K160. Activity of these different RAG1 single chain constructs was monitored using the previously described yeast screening assay (see example 1).

1) Material and Methods

a) Cloning of truncated versions of the M2 mutant in the yeast expression vector

Cloning of the RAG1 M2-RM2-M3 single chain requires first to get the M2 mutant cloned in the yeast expression vector (pCLS0542) (see Material and Methods of Example 3). To clone truncated versions of the M2 mutant, several PCR reactions were performed with different primers couples: CreNter6 / CreCter, CreNter / CreCter160, CreNter / CreCter156, CreNter / CreCter152, CreNter / CreCter145. Sequences of the different primers are listed in the table II below. The different PCR fragments were then digested with the restriction enzymes *Nco*I and *Eag*I, and ligated

into the pCLS0542 vector also digested with NcoI and *Eag*I. The clones were then sequenced. The truncated versions of the M2 mutant are respectively: M2 (6-163), M2(1-160), M2(1-156), M2(1-152) and M2(1-145), where the numbers indicate the I-*Cre*I residues contained in the M2 mutant coding sequence.

b) Cloning of a RAG1 single chain molecule with different endings for both subunits

Different PCR reactions were performed on the M3 mutant in the I-*Cre*I CLS version. Each PCR reaction uses one forward primer and one reverse primer. There are two possible forward primers (RM2 and RM2N2) and five possible reverse primers (CreCterR60, Cre160R60, Cre156R60, Cre152R60 and Cre145R60). The forward primers allow to obtain a M3 coding sequence beginning at residue N6 or N2 and the reverse primers allow to obtain a M3 coding sequence ending respectively at residues P163, K160, S156, L152 and S145. The different PCR fragments were purified and digested by *Eag*I and *Sac*I and each PCR fragment was ligated into the yeast expression vector for one of the M2 mutants described above also digested with *Eag*I and *Sac*I. After sequencing of the clones, all possible single chain molecules in the yeast expression vector were obtained.

Table II: Primers sequences

Primer Name	Sequence (SEQ ID NO: 79 to 92)
CreNter	5'- acaggccatggccaataccaaatataacaaag -3'
CreNter6	5'- acaggccatggccaacaaagagttcctgctgtacctg-3'
CreCter	5'- gattgacggccgccggggaggatttcttctt-3'
CreCter160	5'- gattgacggccgctttcttcttctcgctcaggctgtc -3'
CreCter156	5'- gattgacggccgcgctcaggctgtccaggacagcacg -3'
CreCter152	5'- gattgacggccgccaggacagcacgaacggtttcaga -3'
CreCter145	5'- gattgacggccgcagaagtggttttacgcgtcttaga -3'
RM2	5'- tatcggccggtggatctgataagtataatcaggctctgtctaaatacaaccaagcactgtccaa gtacaatcaggccctgtctggtggaggcggttccaacaaagagttcctgctgtatcttgctggattt -3'
RM2N2	5'- tatcggccggtggatctgataagtataatcaggctctgtctaaatacaaccaagcactgtccaagtaca atcaggccctgtctggtggaggcggttccaacaccaagtacaacaaagagttcctgctgtat -3'
CreCterR60	5'- tagacgagctcctaaggagaggacttttttcttctcag -3'
Cre160R60	5'- tagacgagctcctacttttttcttctcagagaggtcatc -3'
Cre156R60	5'- tagacgagctcctaagagaggtcatccagaactgccct -3'
Cre152R60	5'- tagacgagctcctacagaactgccctcacagtctcaga -3'
Cre145R60	5'- tagacgagctcctaagaggtggtttttctggtcttgga -3'

CLAIMS

1°) A single-chain I-*Cre*I meganuclease comprising two domains joined by a peptidic linker, wherein:

(a) each domain, derived from a parent I-*Cre*I monomer, comprises a portion of said parent I-*Cre*I monomer which extends at least from the beginning of the first alpha helix (α_1) to the end of the C-terminal loop of I-*Cre*I and includes successively: the $\alpha_1\beta_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ core domain, the α_4 and α_5 helices and the C-terminal loop of I-*Cre*I, and

(b) the two domains are joined by a peptidic linker which allows said two domains to fold as a I-*Cre*I dimer that is able to bind and cleave a chimeric DNA target comprising one different half of each parent homodimeric I-*Cre*I meganuclease target sequence.

2°) The single-chain I-*Cre*I meganuclease according to claim 1, wherein the N-terminal domain starts at position 1 or 6 of I-*Cre*I and the C-terminal domain starts at position 2 or 6 of I-*Cre*I.

3°) The single-chain I-*Cre*I meganuclease according to claim 1 or claim 2, wherein the N-terminal and/or C-terminal domain(s) terminate(s) at position 145 of I-*Cre*I.

4°) The single-chain I-*Cre*I meganuclease according to claim 1 or claim 2, wherein the N-terminal and/or C-terminal domain(s) further include(s) at least the alpha 6 helix of I-*Cre*I.

5°) The single-chain I-*Cre*I meganuclease according to claim 4, wherein the N-terminal and/or C-terminal domain(s) terminate(s) at position 152, 156, 160 or 163 of I-*Cre*I.

6°) The single-chain I-*Cre*I meganuclease according to anyone of claims 1 to 5, wherein the linker consists of a sequence of 15 to 35 amino acids.

7°) The single-chain I-*Cre*I meganuclease according to claim 6, wherein the linker is selected from the group consisting of the sequences SEQ ID NO: 2 to 12 and 14 to 19.

8°) The single-chain I-*Cre*I meganuclease according to anyone of claims 1 to 7, wherein both domains comprise different mutations at positions 26 to 40 and/or 44 to 77 of I-*Cre*I, said single-chain I-*Cre*I meganuclease being able to cleave a

non-palindromic DNA sequence, wherein at least the nucleotides at positions +3 to +5, +8 to +10, -10 to -8 and -5 to -3 of said DNA sequence correspond to the nucleotides at positions +3 to +5, +8 to +10 , -10 to -8 and -5 to -3 of a DNA target from a gene of interest.

9°) The single-chain I-*Cre*I meganuclease according to anyone of claims 1 to 8, wherein at least one domain comprises a mutation at positions 137 to 143 of I-*Cre*I that modifies the specificity of the single-chain I-*Cre*I meganuclease towards the nucleotides at positions ± 1 to 2, 6 to 7 and/or 11 to 12 of the I-*Cre*I site.

10°) The single-chain I-*Cre*I meganuclease according to claim 8 or claim 9, wherein said mutations 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 and W.

11°) The single-chain I-*Cre*I meganuclease according to anyone of claims 1 to 10, which comprises mutation(s) that impair(s) the formation of functional homodimers from the two domains.

12°) The single-chain I-*Cre*I meganuclease according to claim 11, wherein each domain comprises at least one mutation, selected from the group consisting of: K7E or K7D and E8K or E8R; F54G or F54A and L97F or L97W; K96D or K96E and E61R or E61K; R51D or R51E and D137R or D137K, respectively for the first and the second domain.

13°) The single-chain I-*Cre*I meganuclease according to claim 12, wherein one domain comprises the substitution of the lysine residues at positions 7 and 96 by an acidic amino acid and the other domain comprises the substitution of the glutamic acid residues at positions 8 and 61 by a basic amino acid.

14°) The single-chain I-*Cre*I meganuclease according to anyone of claims 1 to 10, wherein one domain comprises the G19S mutation.

15°) The single-chain I-*Cre*I meganuclease according to anyone of claims 14, wherein the other domain comprises at least one mutation that impairs the formation of a functional homodimer as defined in anyone of claims 11 to 13.

16°) A polynucleotide fragment encoding the single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15.

17°) The polynucleotide fragment according to claim 16, wherein the nucleotide sequences encoding the two domains have less than 80 % nucleic sequence identity.

18°) The polynucleotide fragment according to claim 17, wherein the sequences encoding the two domains are derived from the sequences SEQ ID NO: 22 and SEQ ID NO: 23, respectively.

19°) An expression vector comprising a polynucleotide fragment of anyone of claims 16 to 18, operatively linked to regulatory sequences allowing the production of the single-chain I-*Cre*I meganuclease.

20°) The vector of claim 19, which includes a targeting DNA construct comprising sequences sharing homologies with the region surrounding a genomic DNA target of said single-chain I-*Cre*I meganuclease.

21°) The vector of claim 20, wherein said targeting DNA construct comprises : a) sequences sharing homologies with the region surrounding said genomic DNA target, and b) sequences to be introduced flanked by sequence as in a).

22°) A host cell comprising a polynucleotide fragment of anyone of claims 16 to 18 or a vector of anyone of claims 19 to 21.

23°) A non-human transgenic animal comprising a polynucleotide fragment of anyone of claims 16 to 18.

24°) A transgenic plant comprising a polynucleotide fragment of anyone of claims 16 to 18.

25°) A pharmaceutical composition comprising at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, or one vector of anyone of claims 19 to 21.

26°) The composition of claim 25, further comprising a targeting DNA construct comprising the sequence which repairs a genomic site of interest flanked by sequences sharing homologies with said genomic site.

27°) Use of at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, one vector of anyone of claims 19 to 21, one host cell of claim 22, one transgenic plant of claim 24, one non-human transgenic mammal of claim 23, for molecular biology, for

in vivo or *in vitro* genetic engineering, and for in *vivo* or *in vitro* genome engineering, for non therapeutic purposes.

28°) Use of at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, one vector of anyone of claims 19 to 21, for the preparation of a medicament for preventing, improving or curing a genetic disease in an individual in need thereof.

29°) Use of at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, one vector of anyone of claims 19 to 21, for the preparation of a medicament for preventing, improving or curing a disease caused by an infectious agent that presents a DNA intermediate in an individual in need thereof .

30°) Use of at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, one vector of anyone of claims 19 to 21, *in vitro*, for inhibiting the propagation, inactivating or deleting an infectious agent that presents a DNA intermediate, in biological derived products or products intended for biological uses or for disinfecting an object.

31°) The use of claim 29 or claim 30, wherein said infectious agent is a virus.

32°) The use of anyone of claims 27 to 31, wherein said meganuclease, polynucleotide, vector, cell, transgenic plant or non-human transgenic mammal are associated with a targeting DNA construct as defined in anyone of claims 20, 21 and 26.

33°) Use of at least one single-chain I-*Cre*I meganuclease of anyone of claims 1 to 15, one polynucleotide fragment of anyone of claims 16 to 18, one vector of anyone of claims 19 to 21, as a scaffold for engineering other meganucleases.

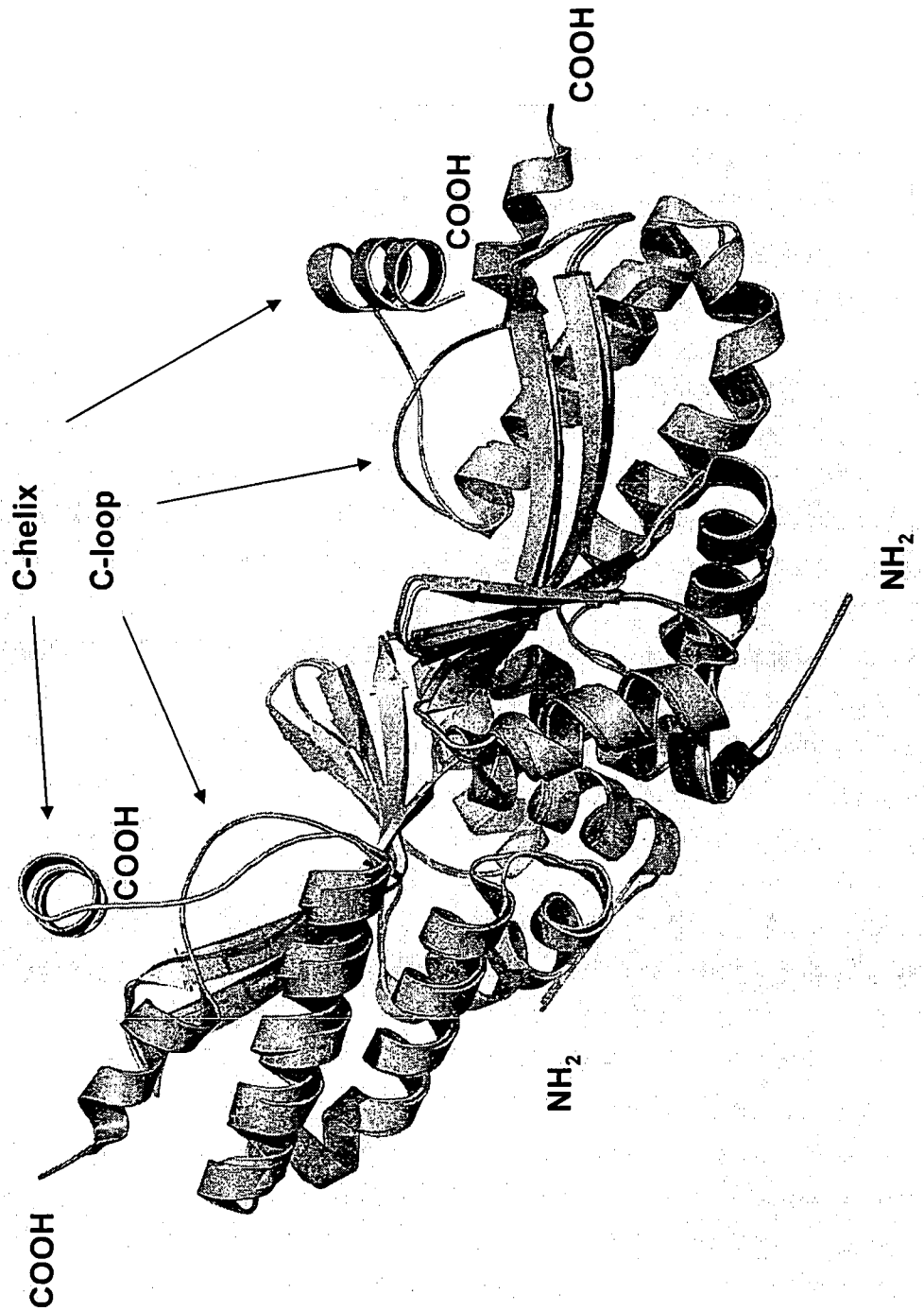

FIGURE 1

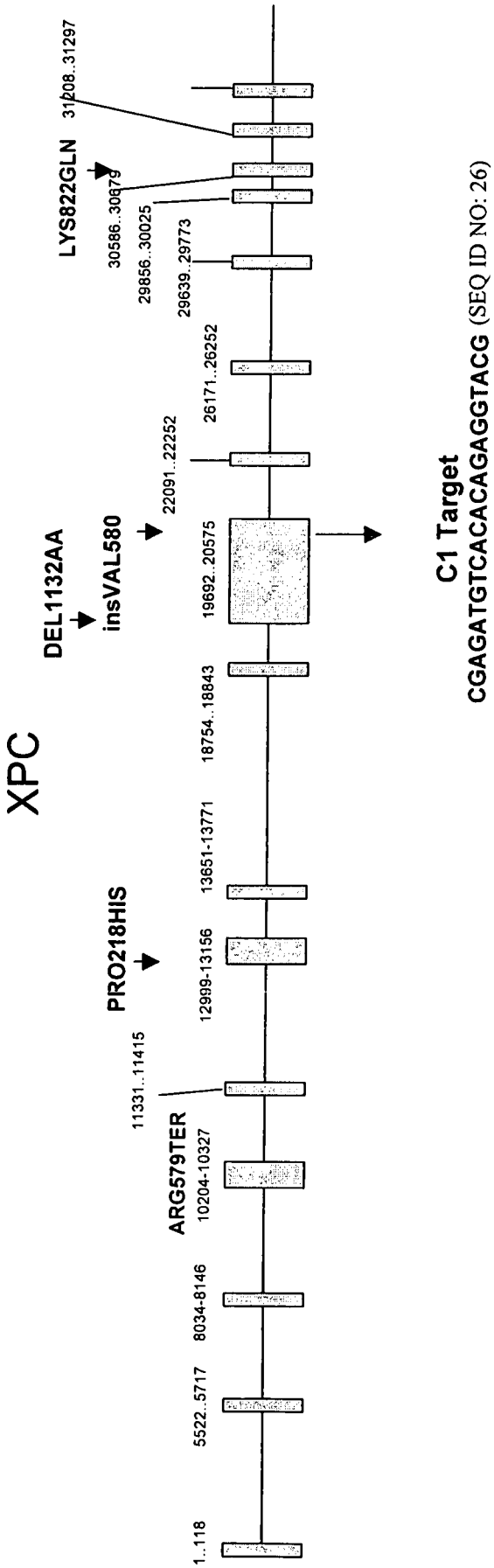
XPC
1..118
5522..5717
8034-8146
10204-10327
11331..11415
12999-13156
13651-13771
18754..18843
19692..20575
22091..22252
26171..26252
29639..29773
29856..30025
30586..30679
31208..31297
ARG579TER
PRO218HIS
DEL1132AA
insVAL580
LYS822GLN
C1 Target
CGAGATGTCACACAGAGGTACG (SEQ ID NO: 26)

FIGURE 2

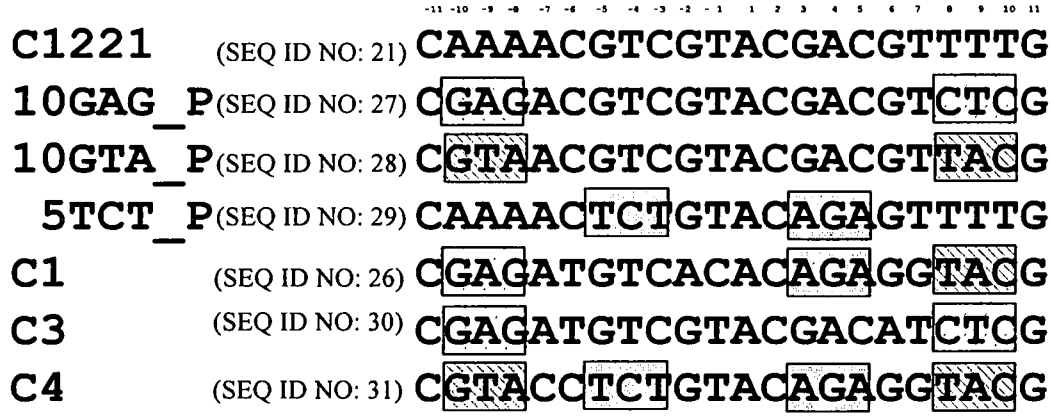
-11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11
C1221 (SEQ ID NO: 21) CAAAACGTCGTACGACGTTTTG
10GAG_P (SEQ ID NO: 27) CGAGACGTCGTACGACGTCTCG
10GTA_P (SEQ ID NO: 28) CGTAACGTCGTACGACGTTACG
5TCT_P (SEQ ID NO: 29) CAAAACTCTGTACAGAGTTTTG
C1 (SEQ ID NO: 26) CGAGATGTCACACAGAGGTACG
C3 (SEQ ID NO: 30) CGAGATGTCGTACGACATCTCG
C4 (SEQ ID NO: 31) CGTACCTCTGTACAGAGGTACG

FIGURE 3

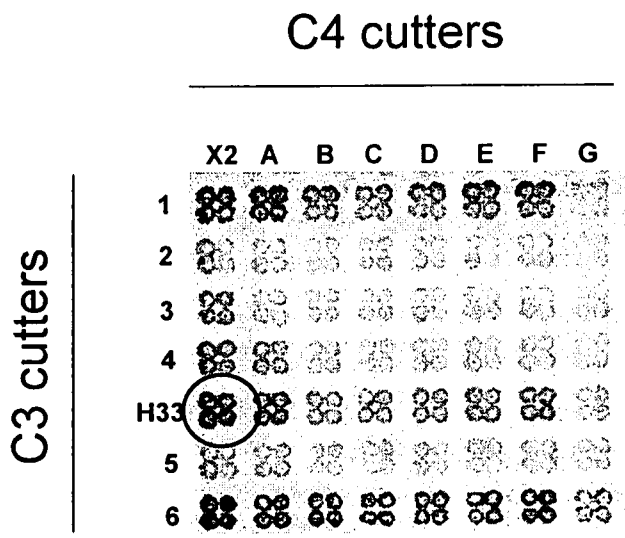

FIGURE 4

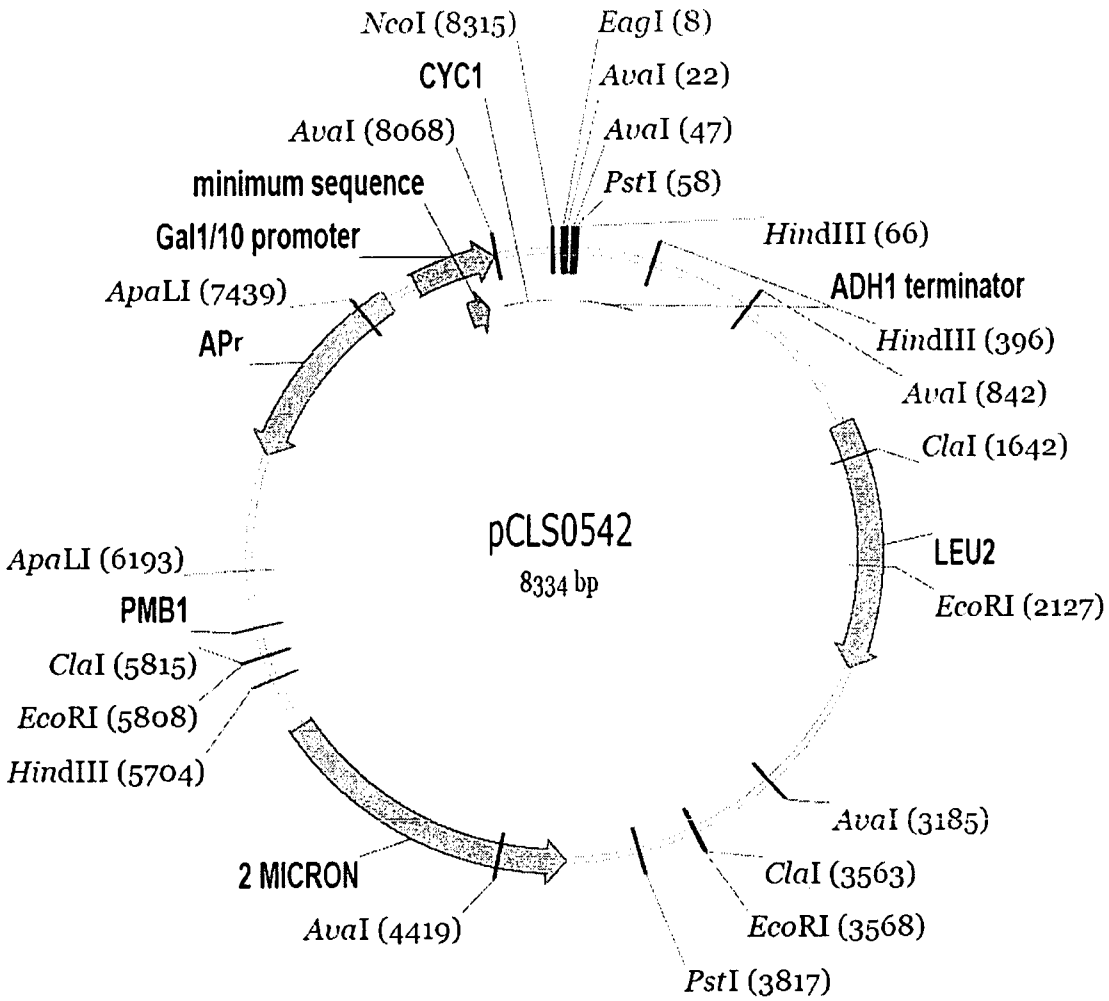
NcoI (8315)
EagI (8)
CYC1
AvaI (22)
AvaI (8068)
AvaI (47)
minimum sequence
PstI (58)
Gal1/10 promoter
HindIII (66)
ApaLI (7439)
ADH1 terminator
APr
HindIII (396)
AvaI (842)
ClaI (1642)
pCLS0542
8334 bp
LEU2
ApaLI (6193)
EcoRI (2127)
PMB1
ClaI (5815)
EcoRI (5808)
AvaI (3185)
HindIII (5704)
ClaI (3563)
2 MICRON
EcoRI (3568)
AvaI (4419)
PstI (3817)

FIGURE 5

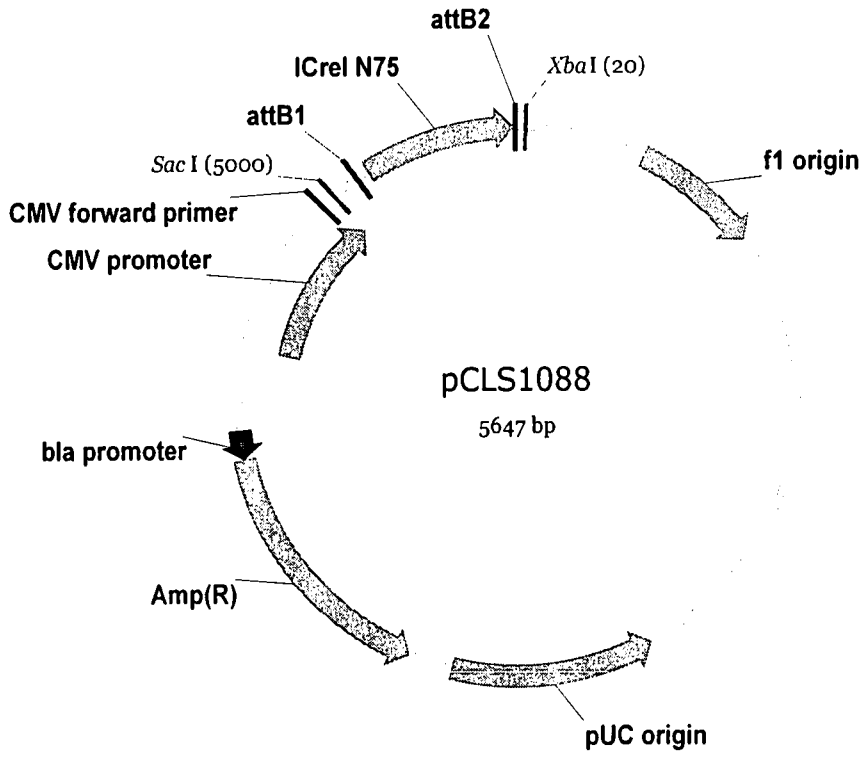
attB2
ICrel N75
XbaI (20)
attB1
SacI (5000)
f1 origin
CMV forward primer
CMV promoter
pCLS1088
5647 bp
bla promoter
Amp(R)
pUC origin

FIGURE 6

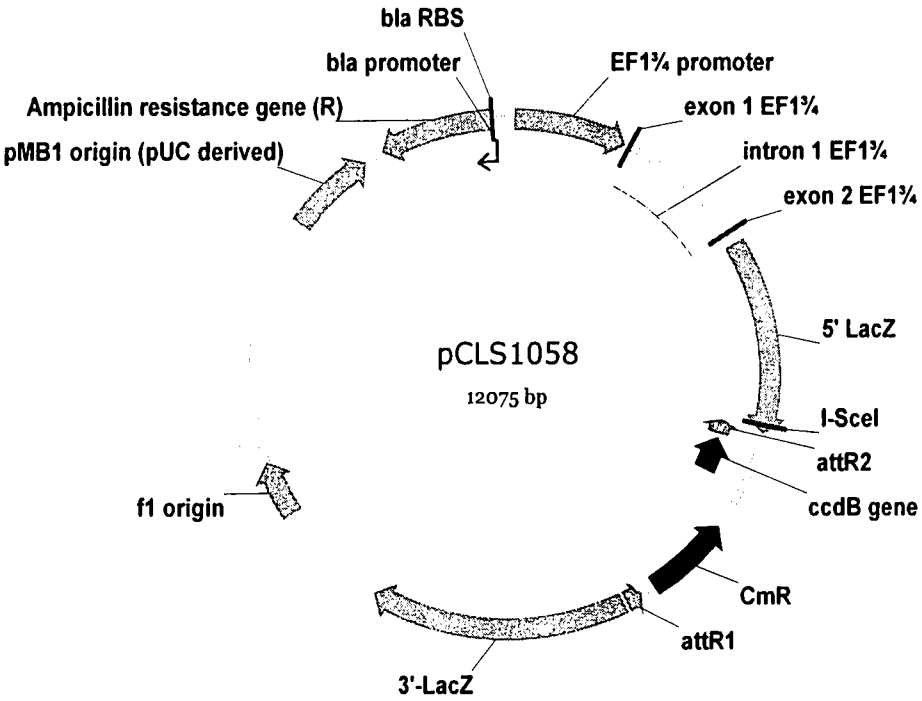
bla RBS
bla promoter
Ampicillin resistance gene (R)
pMB1 origin (pUC derived)
EF1¾ promoter
exon 1 EF1¾
intron 1 EF1¾
exon 2 EF1¾
5' LacZ
pCLS1058
12075 bp
I-SceI
attR2
ccdB gene
f1 origin
CmR
attR1
3'-LacZ

FIGURE 7

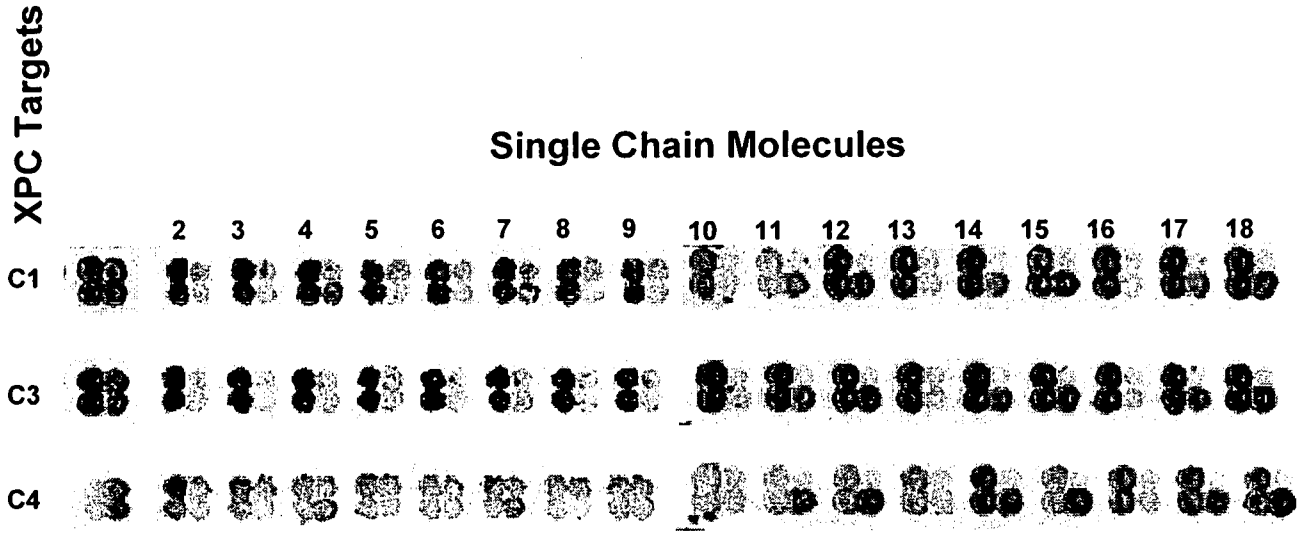
XPC Targets
Single Chain Molecules
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
C1
C3
C4

FIGURE 8

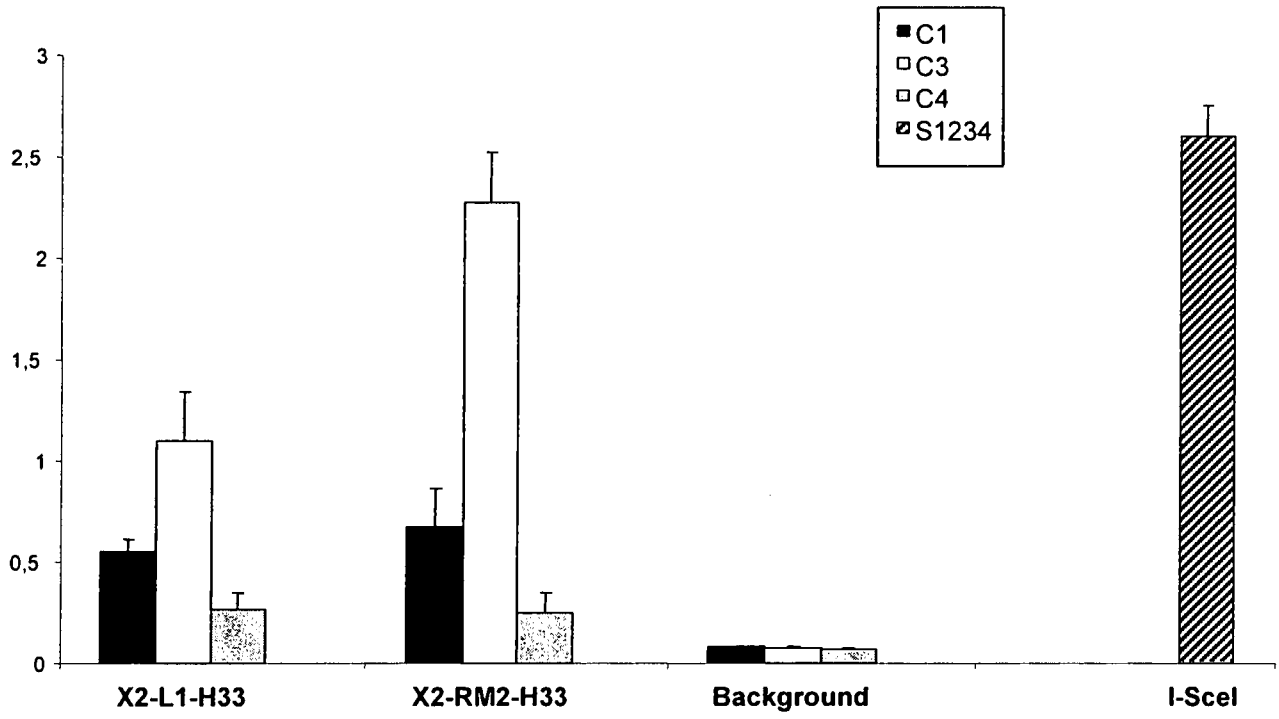
C1
C3
C4
S1234
3
2,5
2
1,5
1
0,5
0
X2-L1-H33
X2-RM2-H33
Background
I-SceI

FIGURE 9

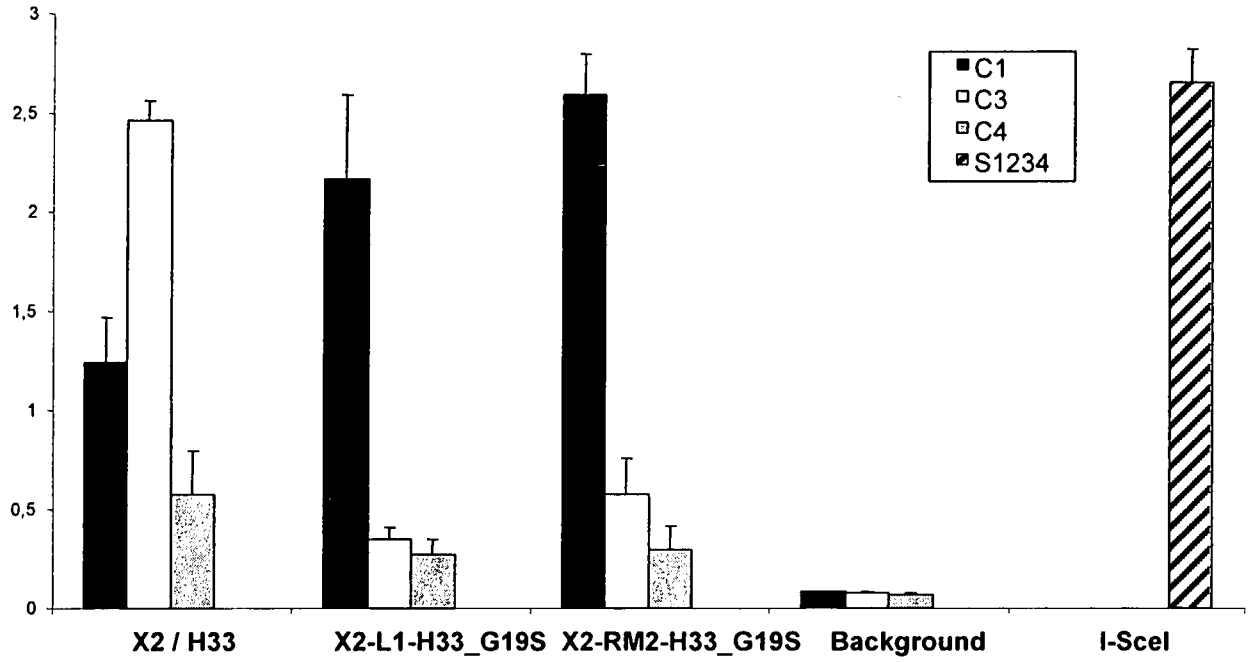
3
2,5
2
1,5
1
0,5
0
C1
C3
C4
S1234
X2 / H33
X2-L1-H33_G19S
X2-RM2-H33_G19S
Background
I-SceI

FIGURE 10

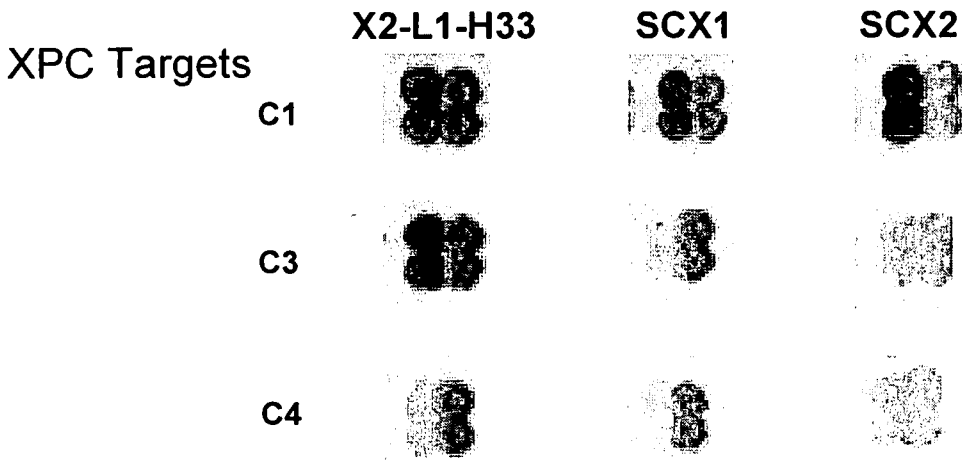

FIGURE 11

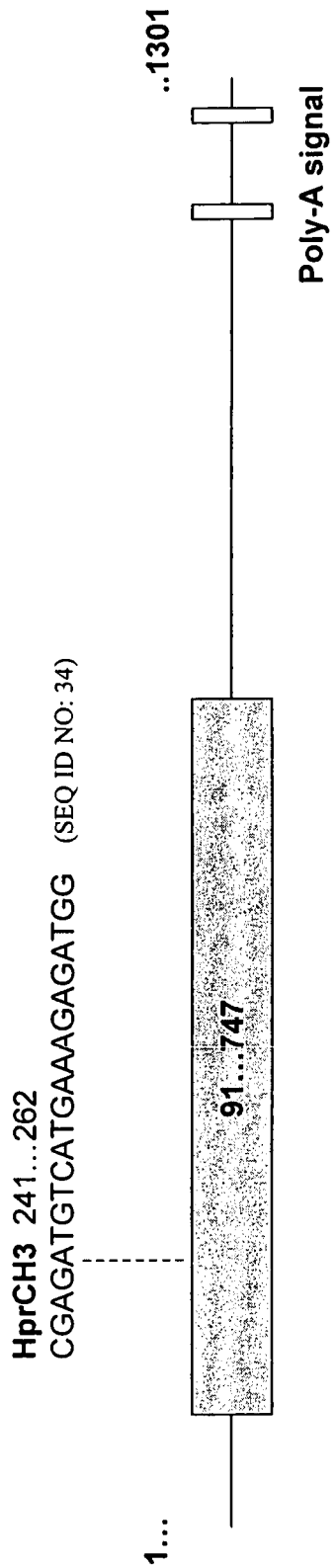
..1301
Poly-A signal
HprCH3 241...262
CGAGATGTCATGAAAGAGATGG (SEQ ID NO: 34)
91...747
1...

FIGURE 12

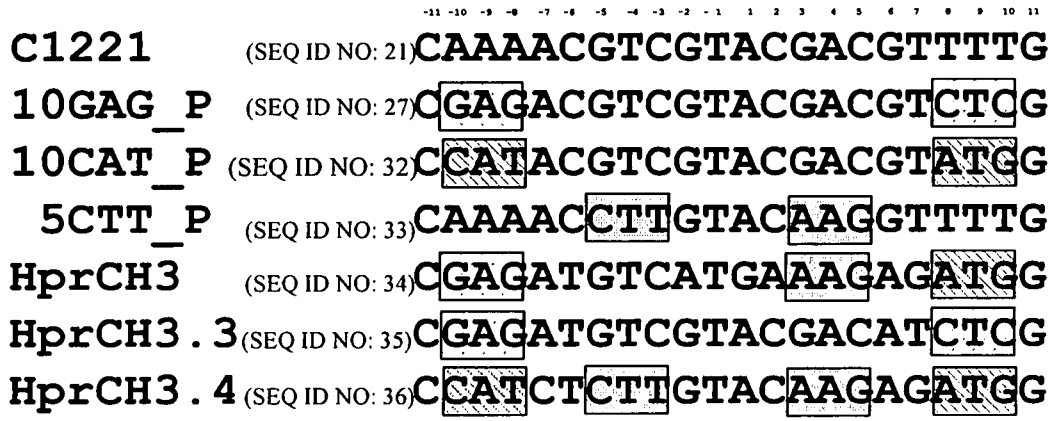
-11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11
C1221 (SEQ ID NO: 21) CAAAACGTCGTACGACGTTTTG
10GAG_P (SEQ ID NO: 27) CGAGACGTCGTACGACGTCTCG
10CAT_P (SEQ ID NO: 32) CCATACGTCGTACGACGTATGG
5CTT_P (SEQ ID NO: 33) CAAAACCTTGTACAAGGTTTTG
HprCH3 (SEQ ID NO: 34) CGAGATGTCATGAAAGAGATGG
HprCH3.3 (SEQ ID NO: 35) CGAGATGTCGTACGACATCTCG
HprCH3.4 (SEQ ID NO: 36) CCATCTCTTGTACAAGAGATGG

FIGURE 13

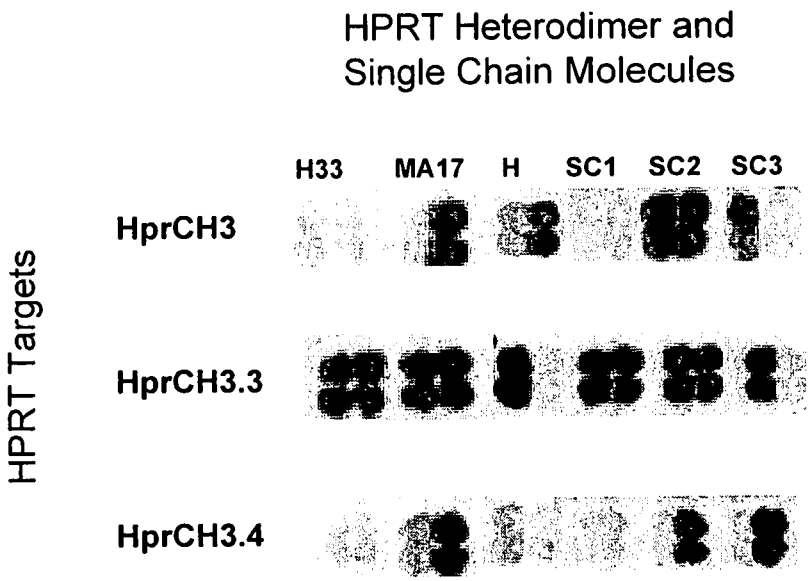
HPRT Heterodimer and
Single Chain Molecules
H33
MA17
H
SC1
SC2
SC3
HprCH3
HprCH3.3
HprCH3.4
HPRT Targets

FIGURE 14

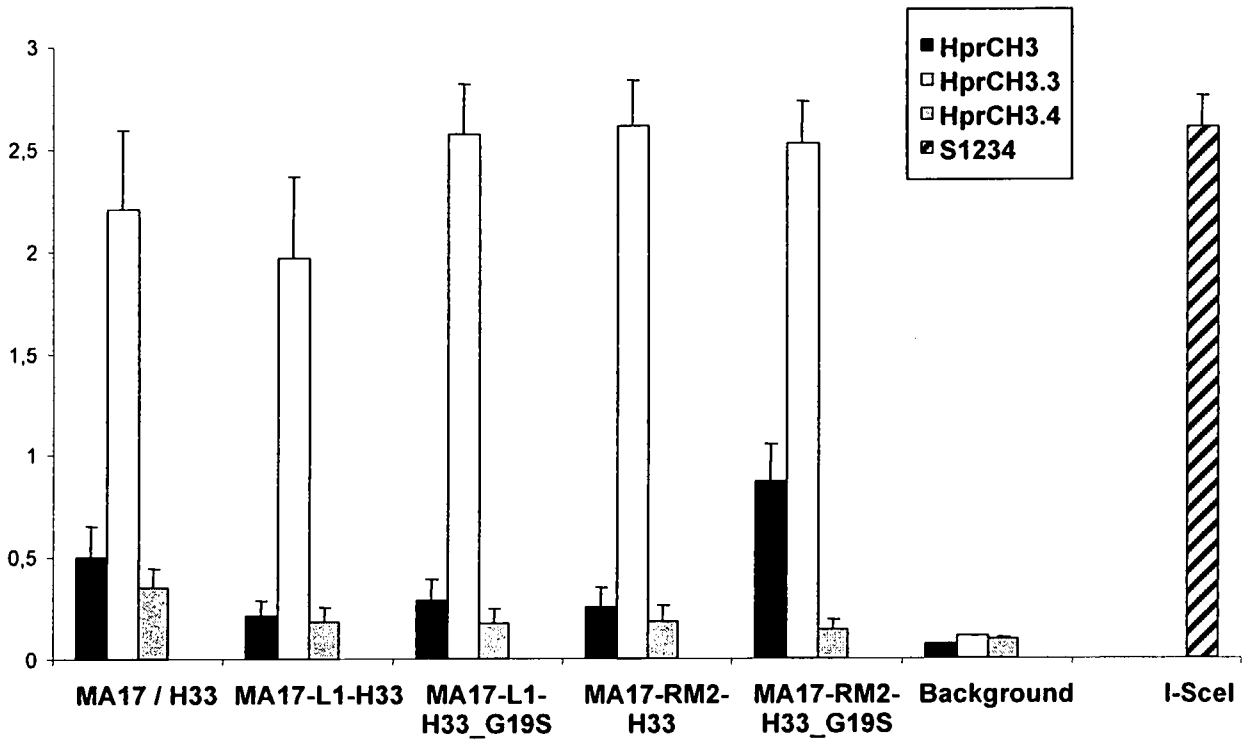
HprCH3
HprCH3.3
HprCH3.4
S1234
3
2,5
2
1,5
1
0,5
0
MA17 / H33
MA17-L1-H33
MA17-L1-H33_G19S
MA17-RM2-H33
MA17-RM2-H33_G19S
Background
I-SceI

FIGURE 15

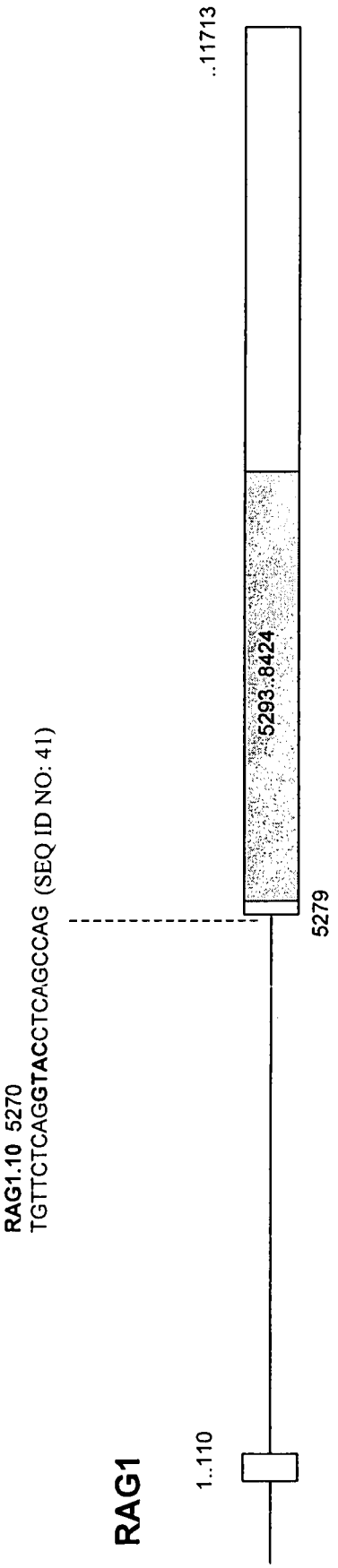
RAG1
1..110
..11713
5293..8424
5279
RAG1.10 5270
TGTTCTCAGGTACCTCAGCCAG (SEQ ID NO: 41)

FIGURE 16

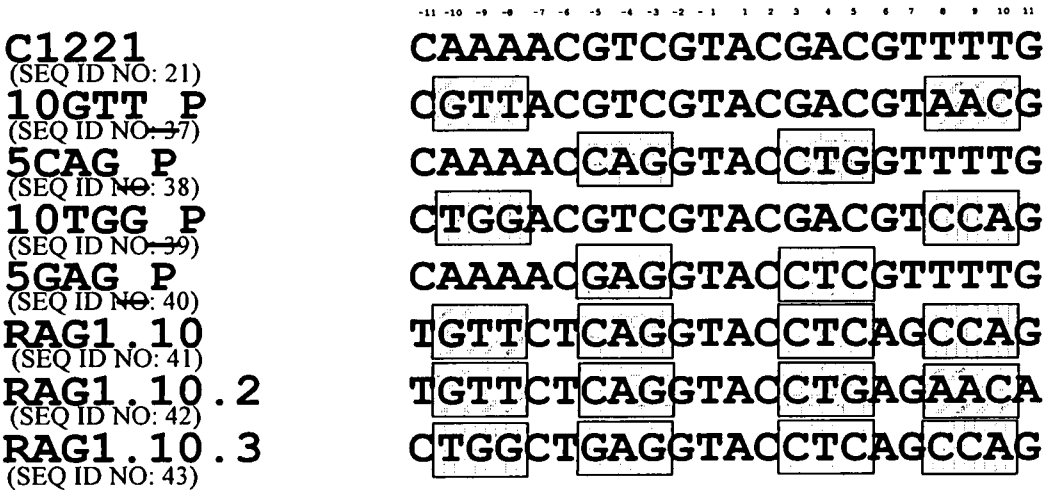
-11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11
C1221
(SEQ ID NO: 21)
CAAAACGTCGTACGACGTTTTG
10GTT_P
(SEQ ID NO: 37)
CGTTACGTCGTACGACGTAACG
5CAG_P
(SEQ ID NO: 38)
CAAAACCAGGTACCTGGTTTTG
10TGG_P
(SEQ ID NO: 39)
CTGGACGTCGTACGACGTCCAG
5GAG_P
(SEQ ID NO: 40)
CAAAACGAGGTACCTCGTTTTG
RAG1.10
(SEQ ID NO: 41)
TGTTCTCAGGTACCTCAGCCAG
RAG1.10.2
(SEQ ID NO: 42)
TGTTCTCAGGTACCTGAGAACA
RAG1.10.3
(SEQ ID NO: 43)
CTGGCTGAGGTACCTCAGCCAG

FIGURE 17

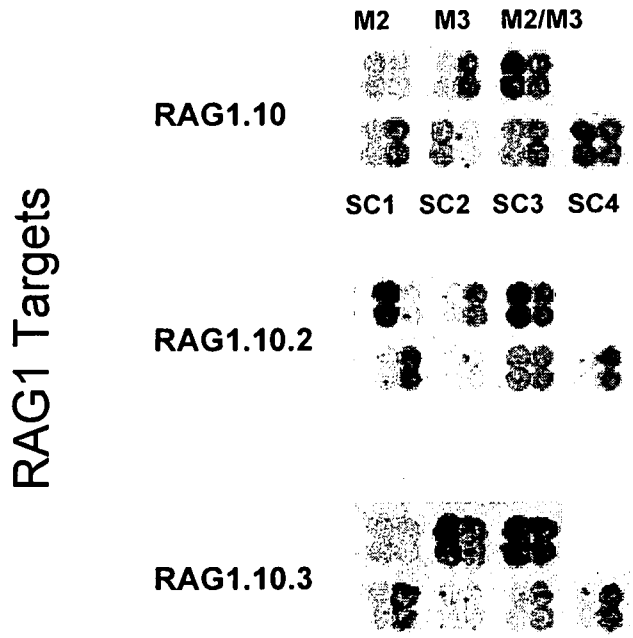
RAG1 Targets
M2
M3
M2/M3
RAG1.10
SC1
SC2
SC3
SC4
RAG1.10.2
RAG1.10.3

FIGURE 18

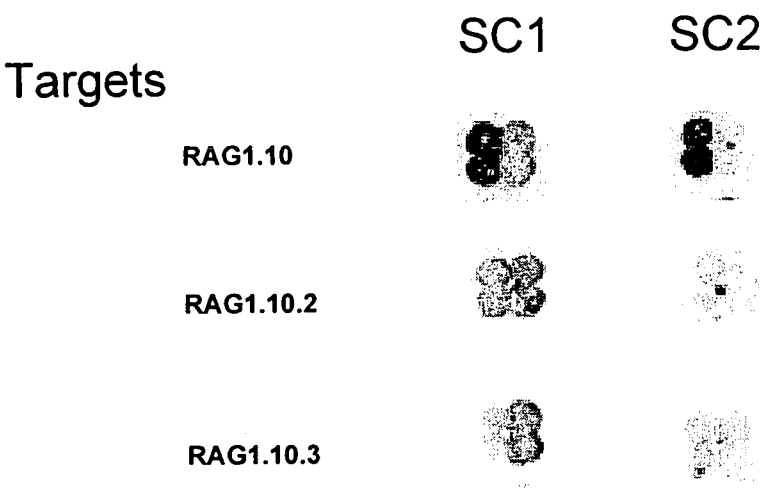

FIGURE 19

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2008/001331

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, EMBASE, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/078619 A (CELLECTIS [FR]; ARNOULD SYLVAIN [FR]; CHAMES PATRICK [FR]; CHOULIKA AN) 25 September 2003 (2003-09-25) cited in the application	1,16,19, 22,27
Y	the whole document	1-33

Y	EPINAT JEAN-CHARLES ET AL: "A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells." NUCLEIC ACIDS RESEARCH, vol. 31, no. 11, 1 June 2003 (2003-06-01), pages 2952-2962, XP002248751 ISSN: 0305-1048 cited in the application the whole document	1-33

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 November 2008	21/11/2008

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Valcárcel, Rafael

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2008/001331

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PRIETO JESUS ET AL: "The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage" NUCLEIC ACIDS RESEARCH, vol. 35, no. 10, May 2007 (2007-05), pages 3262-3271, XP009108728 ISSN: 0305-1048 cited in the application the whole document	1-33
Y	SILVA ET AL: "From Monomeric to Homodimeric Endonucleases and Back: Engineering Novel Specificity of LAGLIDADG Enzymes" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB, vol. 361, no. 4, 25 August 2006 (2006-08-25), pages 744-754, XP005593236 ISSN: 0022-2836 the whole document	1-33
Y	ARNOULD ET AL: "Engineered I-CreI Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB, vol. 371, no. 1, 9 July 2007 (2007-07-09), pages 49-65, XP022145891 ISSN: 0022-2836 cited in the application the whole document	1-33
X	WO 2006/097853 A (CELLECTIS [FR]; DUCHATEAU PHILIPPE [FR]; PAQUES FREDERIC [FR]) 21 September 2006 (2006-09-21)	1,16,19, 22
Y	the whole document	1-33
Y	SMITH JULIANNE ET AL: "A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences" NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, vol. 34, no. 22, 27 November 2006 (2006-11-27), pages e149-1, XP002457876 ISSN: 0305-1048 cited in the application the whole document	1-33

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2008/001331

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHEVALIER B S ET AL: "Design, activity, and structure of a highly specific artificial endonuclease" MOLECULAR CELL, CELL PRESS, CAMBRIDGE, MA, US, vol. 10, no. 4, 20 October 2002 (2002-10-20), pages 895-905, XP002248750 ISSN: 1097-2765 cited in the application the whole document	1-33
X	WO 2007/049095 A (CELLECTIS [FR]; PAQUES FREDERIC [FR]) 3 May 2007 (2007-05-03)	1,16,19, 22
Y	the whole document	1-33
X	WO 2007/093918 A (CELLECTIS [FR]; ARNOULD SYLVAIN [FR]; PEREZ-MICHAUT CHRISTOPHE [FR]; S) 23 August 2007 (2007-08-23) cited in the application	1,16,19, 22
Y	the whole document	1-33
X	WO 2006/097784 A (CELLECTIS [FR]; PAQUES FREDERIC [FR]) 21 September 2006 (2006-09-21)	1,16,19, 22
Y	the whole document	1-33
Y	WO 2007/093836 A (CELLECTIS [FR]; PEREZ-MICHAUT CHRISTOPHE [FR]) 23 August 2007 (2007-08-23) cited in the application the whole document	1-33
E	WO 2008/093249 A (CELLECTIS [FR]; FAJARDO SANCHEZ EMMANUEL [ES]; GRIZOT SYLVESTRE [FR];) 7 August 2008 (2008-08-07) the whole document	1,16,19, 22

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2008/001331

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03078619 A	25-09-2003	AT 372379 T	15-09-2007
		AU 2003215869 A1	29-09-2003
		CA 2479153 A1	25-09-2003
		DE 60316124 T2	05-06-2008
		DK 1485475 T3	21-01-2008
		EP 1485475 A1	15-12-2004
		ES 2292994 T3	16-03-2008
		JP 2005520519 T	14-07-2005
WO 2006097853 A	21-09-2006	AU 2006224248 A1	21-09-2006
		CA 2600033 A1	21-09-2006
		EP 1863909 A1	12-12-2007
		WO 2006097854 A1	21-09-2006
		JP 2008535484 T	04-09-2008
WO 2007049095 A	03-05-2007	AU 2006314220 A1	24-05-2007
		CA 2625774 A1	24-05-2007
		EP 1979478 A2	15-10-2008
		WO 2007057781 A2	24-05-2007
WO 2007093918 A	23-08-2007	AU 2007216264 A1	23-08-2007
		CA 2639149 A1	23-08-2007
		EP 1989299 A2	12-11-2008
		WO 2007093836 A1	23-08-2007
WO 2006097784 A	21-09-2006	CN 101198694 A	11-06-2008
WO 2007093836 A	23-08-2007	AU 2007216264 A1	23-08-2007
		CA 2639149 A1	23-08-2007
		EP 1989299 A2	12-11-2008
		WO 2007093918 A2	23-08-2007
WO 2008093249 A	07-08-2008	WO 2008093152 A1	07-08-2008